

223  
27.8.70

The  
British Journal  
of  
Educational Psychology  
(Incorporating the "Forum of Education")

EDITED BY  
P. E. VERNON

AIDED IN THE SELECTION OF PAPERS BY

SIR CYRIL BURT  
R. A. C. OLIVER  
E. A. PEEL

C. W. VALENTINE  
S. WISEMAN  
HELEN M. WODEHOUSE

AND THE FOLLOWING REPRESENTATIVES  
OF OTHER COUNTRIES:

L. CARMICHAEL  
ARNOLD GESELL  
I. LORGE

I. MEYERSON  
JEAN PIAGET  
FRED J. SCHONELL

ASSISTANT EDITOR  
L. B. BIRCH

Volume XXX  
1960



Published by  
METHUEN AND CO., LTD., 36, ESSEX STREET, STRAND, LONDON, W.C.2,  
for the British Journal of Educational Psychology, Ltd.  
Registered Offices: 42, Hill Street, Birmingham, 5.  
on behalf of

THE BRITISH PSYCHOLOGICAL SOCIETY and THE ASSOCIATION  
OF TEACHERS IN COLLEGES AND DEPARTMENTS OF EDUCATION

*All rights reserved.*

Bureau Ednl. Res. Research  
DAVID H. 2. - COLLEGE  
Dated... 27. 8. 70  
Iss. No. 8 223





AGGRESSIVE AND WITHDRAWING CHILDREN IN THE  
NORMAL SCHOOL

## I—PATTERNS OF BEHAVIOUR.

By E. A. LUNZER

*(Department of Education, University of Manchester.)*

SUMMARY. Forty-two "aggressive" and forty "withdrawing" children, selected by teachers out of a total school population of 1,002 children, are compared with each other and with a group of "exceptionally well-adjusted children." Comparisons are based on rating scales, Bristol Guides to Social Adjustment, a sociometric test, and a test of social adaptation.

The results indicate that there are probably rather more problems of aggression than withdrawal but that the severity of the two behaviour patterns is comparable. These indications are confirmed by a limited one-year follow-up.

## I.—INTRODUCTION.

THE work of Wickman (1928) is one of the best known studies in the field of 'mental hygiene' in school-children. A group of teachers were asked to rank a list of symptoms of maladjustment in what they considered to be their order of severity. A similar list was ranked by a group of psychologists, psychiatrists and psychiatric social workers. Teachers gave a great deal more emphasis to violations of authority and other 'aggressive' forms of behaviour, while those kinds of behaviour which might be held to characterise the shy, recessive personality were relegated to a very low position in their rankings, suggesting that these symptoms were not regarded as serious in their import. The details of Wickman's procedure have been rightly criticised on a number of grounds which need not concern us here. It remains that his investigation has been repeated many times and the results have been substantially the same, even where considerable care was taken to avoid the procedural errors to be found in the original. The whole of this work has been admirably reviewed in a recent and carefully conducted study by Hollins (1955). This investigation involved a comparison between the rankings of Manchester teachers and those of a group of P.S.Ws. In general, the findings of Wickman were again confirmed.

There is an unstated value judgment implicit in such studies: it is tacitly assumed that the opinions of the clinicians are a correct reflection of the facts, and that if teachers' opinions differ from theirs, it is because teachers are insufficiently aware of the essential principles of child development. Is this something more than an *argumentum ex auctoritate*? The implications of childhood maladjustment for adult mental health are still very largely *terra incognita*. Long-term studies at present being carried out by the London Institute will, no doubt, go far to remedy this deficiency. At present, the association between persistent delinquent behaviour in childhood and adult criminality is better established than the connection between recessive behaviour in childhood and dysthymic neuroses in later life. This is not to say that because more is known about the relation between aggressive or delinquent behaviour and adult criminality than about the implications of recessive symptoms in children, therefore the latter are necessarily less grave in their prognosis. It may be that they are indeed more serious. Nevertheless, there appears to be a need for more factual knowledge. The present investigation was undertaken as a step in that direction.



## *Aggressive and Withdrawing Children*

### II.—SAMPLING PROCEDURE.

Some twenty advanced students, teachers attending the course in Child Psychology for the Diploma in Education at Manchester University, were asked to put forward one or two names of children in their class or classes who were most 'aggressive', and one or two, depending on the size of the class, who were most 'withdrawing', together with an equivalent number of 'exceptionally well-adjusted' children. The aim was to obtain two experimental groups who might reasonably be described as deviants at the 5 per cent. level—which would correspond in all to 10 per cent. of the population from which they were taken. That figure is not disproportionate to the percentages reported in the three pilot surveys carried out for the Committee on Maladjusted Children as requiring some psychiatric attention. Each of these divided the children in a representative cross-section into five categories. Category A were those rated as seriously maladjusted, requiring clinical treatment; category B were those who were probably maladjusted and should be recommended for psychiatric interview, but not necessarily for treatment. The remaining three categories represented freedom from psychiatric symptoms, or relatively minor troubles which could be countered without recourse to the child guidance service. Overall percentages in the first two categories combined were 11·3 in Somerset, 15·8 in Birmingham, and 5·4 in Berkshire.

Teachers were encouraged to nominate more than the minimum for the two 'experimental' categories if they were especially unfortunate in the class that they had, but in point of fact, none of them did so, although there are reasonable indications from a consideration of those children who were included in the sample that some of them might have done so. On the other hand, several teachers put forward fewer names than their statistical allocation, on the ground that they had not enough children in their classes who might be regarded as even remotely 'unsettled', and they may well have been justified.

Since not all the children who were nominated were eventually rated and tested in the course of the subsequent investigation, the final samples were a little smaller than the original, and differed, albeit slightly, from one comparison to another. For most purposes the three samples consisted of forty-five 'well-adjusted' children, forty-two 'aggressive' and forty 'withdrawing'. Only three children in each group were in secondary modern schools, and of the remainder, about ten in each group were infants and the others were of junior age. The total population from which these selections were made consisted of twenty-eight classes drawn from thirteen schools in and around Manchester, i.e., 1,002 children in all.

TABLE 1

AGES OF CHILDREN REPRESENTED IN THE THREE SAMPLES.

(W=Withdrawing. WA=Well-adjusted. A=Aggressive.)

Age-Group	W	WA	A	Total
5 — 7·55	8	9	8	25
7·6— 9·6	12	13	13	38
9·6—11·5	17	20	18	55
11·6 plus	3	3	3	9

Analysis of variance showed the three samples to be strictly comparable in respect of age.

The teachers who made the selection were free to nominate boys or girls, whichever were best fitted to the category for which the selections were made. With a single exception, all of the twenty-eight classes concerned were mixed. In the event the well-adjusted sample consisted of twenty-one boys and twenty-four girls, the withdrawing sample of eighteen boys and twenty-two girls and the aggressive sample of thirty-three boys and only nine girls. Sex differences in the first two groups are not significant, but in the aggressive group, with  $\chi^2=6.3$ , the difference is significant at the .02 level. This fits in with the observation that more boys than girls are referred to child guidance clinics. Since the incidence of recessive behaviour is approximately equal in the two sexes, while aggressive behaviour is considerably more prevalent in boys, and since referrals by parents, teachers and juvenile courts are more likely to be on the grounds of 'difficult' behaviour than excessive shyness, as we know from Wickman, Stogdill (1931) and their followers, it is not surprising that referrals as a whole, that is including those made by L.E.As. and G.Ps., should number more boys than girls. It also suggests that aggressive behaviour is more frequent as a whole than recessive—although it tells us nothing as to which is more serious.

Comparisons between the three samples with respect to size of family, position in family and whether the mothers were at home or at work all showed insignificant differences, although the trends (more youngest and only children in the withdrawing sample, more mothers working full-time in the aggressive sample) were in line with expectation.

### III.—RATINGS.

All of the children in each of the three samples were assessed on each of twelve traits. These were selected from a number of available lists\* to cover a fairly wide range of behaviour with a minimum of overlap. Careful precautions were taken to minimise halo effect. All the children in each class concerned were ranked on separate occasions for each trait, and the ratings themselves were arrived at subsequently by allocating all of these to one of the five categories in approximately a normal distribution. It will be seen that children included in these samples differ very widely from the normal distribution which characterised the population from which they were drawn. The number of children assessed at each of the 5 points on each of the twelve traits is shown in Table 2.

\* Schonell (1944), N.F.E.R. (1952) and Fleming (1954).



TABLE 2

ASSESSMENTS ON RATING SCALES.

TRAIT-RATINGS ON A FIVE-POINT SCALE AS MADE BY CLASS TEACHERS.

	1	2	3	4	5	
1. Aggression : Aggressive—Compliant/Timid . . . .	— — 22	— 3 15	6 37 5	17 5 —	17 — —	W WA A
	Highly significant					
2. Concentration : Absorbed—Wandering . . . . .	1 11 2	4 18 9	18 15 10	12 1 10	5 — 11	W WA A
	Not significant.					
3. Self-Confidence : Very confident/cocksure—Over dependent . . . . .	2 8 8	— 7 9	5 27 16	23 3 5	10 — 4	W WA A
	Highly significant.					
4. Persistence in Failure : Purposeful—Depressed/Resentful	1 11 7	3 23 9	16 9 8	12 2 15	8 — 3	W WA A
	Significant at .01 level.					
5. Maintenance of Effort : Sustained enthusiasm—Listless/ Apathetic . . . . .	— 12 2	8 20 10	14 11 12	10 2 17	8 — 1	W WA A
	Not significant.					
6. Sensitiveness to Praise and Blame : Extremely sensitive—Quite indifferent. . . . .	10 2 1	12 6 3	14 35 15	4 2 19	— — 4	W WA A
	Highly significant.					
7. Neat-Handedness : Skilful—Clumsy . . . . .	1 11 5	4 18 5	19 13 19	10 3 12	6 — 1	W WA A
	Not significant.					
8. Attitude to Adult Authority : Highly co-operative—Frequently obstructive/Rebellious . . . . .	1 24 1	4 14 9	30 7 13	5 — 11	— — 8	W WA A
	Highly significant.					
9. Sociability : Only happy with group—Prefers to be alone . . . . .	— 3 2	2 13 20	10 27 11	22 2 8	6 — 1	W WA A
	Highly significant.					
10. Initiative : Marked originality and drive— Very dependent/diffident . . .	— 9 4	— 23 16	7 13 13	19 — 8	14 — 1	W WA A
	Highly significant.					
11. Response in Class : Exceptionally good talker— Hesitant/Difficult to draw out . . . . .	— 10 4	— 13 16	5 20 17	16 2 4	19 — 1	W WA A
	Highly significant.					
12. Conscientiousness : High standard of work/Scrupulous —Irresponsible/No ambition in work . . . . .	— 8 2	9 21 6	19 15 17	10 1 14	2 — 3	W WA A
	Not significant.					



Differences with respect to 'aggression' merely indicate that not all children shared symptoms of aggression or withdrawal to the same degree. The ratings given in respect of concentration, persistence in failure, and maintenance of effort are of particular interest, since these may be held to indicate the effect of maladjustment without being specific as to its direction. It may be said that maladjustment in general is a response to stress characterised by anxiety. The nature of the response will vary from one individual to the next and, in particular, it will be greatly different in the recessive children from the aggressive. But whatever the specific response pattern, there will be inhibition of the highest centres of nervous activity, resulting in a loss of concentration, in lack of persistence and in lack of enthusiasm for work tasks.

It appears that more children in the aggressive group show an extreme inability to concentrate, but the difference between this group and the withdrawing is not significant ( $\chi^2=6.9$ , d.f.=4). Differences with respect to persistence in failure are significant. Pooling categories 1+2 and 4+5 in a  $3 \times 2$  table,  $\chi^2=10.0$  which is significant at the .01 level (figures for the well-adjusted group are not included in any calculations, since these are clearly significantly different, but reveal no more than the general importance of good adjustment in fostering favourable behaviour traits). Again, the distribution of the two groups when assessed for maintenance of effort favours the aggressive. The disproportion between categories 4 (Only fitful application. Fairly frequent rests.) and category 5 (Seems too listless and apathetic to do much) is striking. It is in line with what is known of asthenic neuroses but it accords ill with the characterisation of introverts as readily conditioned or excitatory. Much may depend on the stability of the nervous system and on the level of complexity at which the conditioning is demanded.

Taken as a whole, assessments on the three traits do not warrant more than limited conclusions. There appears to be little difference in the incidence of more or less anxiety in the two groups insofar as concentration on day-to-day tasks is impaired, but the withdrawing group show less determination in the face of failure and a greater tendency to fatigue. However, any inferences as to the quality of work produced by the two groups will be seen to be not warranted by the facts, as is shown in a forthcoming paper (Lunzer, 1960).

The distribution of ratings on the remaining traits point to the marked differences between the two groups of maladjusted children. In several instances the aggressive and withdrawing groups are at opposite extremes, with the well-adjusted control group midway between them. The withdrawing group are significantly more dependent and more sensitive to praise and blame.\* They show less initiative and less response in class. It would appear that they are also less sociable, although this is not confirmed by the index of popularity as shown in sociometric choices made by their classmates. Differences with respect to conscientiousness and neat-handedness are not significant.

#### IV.—BRISTOL GUIDES.

A further indication of the degree of maladjustment present in each of the two experimental groups is provided by an inspection of the Bristol Social Adjustment Guides. These were completed by the class teacher concerned in respect of every child included in the sample. Following on the procedure suggested by Stott (1958) each indicator of maladjustment was given a weighting of two points while the milder symptoms of 'unsettledness' as also the indicators

\* It may be of some significance that the ten withdrawing children rated as extremely sensitive, were all among the youngest in the sample. Do these children develop a thicker skin as they grow older?

of nervousness and various miscellaneous nervous symptoms were scored as one point. The total number of points awarded to each child thus gives some indication of the degree of disturbance that is present, as measured by the tendency for that child to show more or less considerable deviations from the normal in various aspects of his behaviour. The results are shown in the following table:

TABLE 3  
MALADJUSTMENT SCORE ON BRISTOL GUIDES.

	W	WA	A
Mean .....	17.56	2.40	20.33
N .....	39	45	42
S.D. ....	10.46	3.18	12.34

It will be seen that both experimental groups differ very widely from the control group, but there is little difference between the aggressive and the withdrawing. Here it would appear that the former tend to show more disturbance than the latter, but the differences are not significant. It happens that the Bristol Guides tend to be loaded with items indicative of delinquent tendencies, and if there is anything surprising about these results, it is surely that the differences are so small. Analysis of variance indicated that the degree of disturbance consequent on the predominance of one or other pattern of reaction was not affected by age.

From the qualitative point of view, the Bristol Guides yield a far more telling picture of the type of personality disturbance than can be had from this rough and ready method of scoring and it is of some interest to note the extent to which behaviour deviations fall into distinct patterns for the two groups. Here the hypothesis would be that both groups might be expected to show symptoms of restlessness, nervousness, depression and miscellaneous nervous symptoms, but that the withdrawing group would be characterised by signs of unforthcomingness and withdrawal, while the aggressive group would tend to show combinations of anxiety for affection or approval on the one hand, and, on the other, hostility both in relation to adults and in relation to their own peer group. The dichotomy was fairly clear-cut in thirty-one out of the forty-two aggressive children, and twenty-seven out of the forty withdrawing, while the predominance of one pattern over the other was marked in a further five aggressive and five withdrawing. In four out of the forty-two aggressive, and seven out of the forty withdrawing, the ambiguity of behaviour was more marked. It would be unsafe to accept these tendencies as anything more than a tentative indication of the forms taken by disturbance in childhood, since the method of selection acted in such a way as to reduce the number of children in whom the disturbance took a more ambiguous form. However, it is interesting to note that in a recent paper, Stott (1959) has shown that certain items in the Bristol Guides are better predictors of delinquency than others. These items correspond to those symptomatic of HA (hostility to adults), HC (hostility to children) and K ('knavery' or anti-social tendencies), although not all of the items listed under these heads are equally associated with delinquency. In effect, Stott's study finds the same dichotomy between a withdrawing, non-delinquent group and an aggressive, delinquent or potentially delinquent group as that used in the present investigation.\*

\* A similar clear-cut dichotomy emerges in a recent study by Peterson, *et al.* (1959).



In view of the comparable scoring of the two groups when assessed by means of these Guides, they afford some indication of the degree of disturbance present in each case. If an arbitrary figure of 16† is taken as indicative of more or less serious disturbance, then twenty out of the forty withdrawing cases may be said to be 'maladjusted' and twenty-four out of the forty-two aggressive. Altogether, this would be equivalent to forty-four out of the total population of 1,002 or 4.4 per cent. It would appear that the incidence of disturbance as a whole is approximately equally divided in this sample as between the two behaviour patterns. However, in view of the sex differences already noted, it is probable that more cases might have been brought to our notice in the aggressive sample, although, for the most part, these would have been milder. In addition, there is an unknown but probably limited number of children whose names were deliberately withheld because the disturbance was too ambiguous.

#### V.—SOCIOMETRIC CHOICE-RANKING.

An important index of the degree to which the adjustment of a child is impaired by anxiety-induced behavioural patterns may be had from an examination of his capacity to gain the friendship and approval of his peers. Accordingly, a sociometric test was carried out by the teachers of the majority of the classes concerned (over the age of 7). Three criteria were used, with the first three choices weighted 5, 3 and 2. In addition, to allow for an indication of 'social expansiveness' each child was allowed an unlimited number of choices 'below the line' and these were given unit weightings. The total score obtained by each child was found from the class sociomatrix and all the children in the class were then ranked. Scores of children in the three samples are percentile rankings and are, therefore, comparable as between classes of different sizes. These are shown in Table IV.

TABLE 4  
SOCIOMETRIC PERCENTILE RANKINGS.

	W/U	WA	A
Mean .....	37.26	79.61	37.27
N .....	27	31	29
S.D. ....	28.00	19.73	25.89

The differences between the two experimental groups and the control is significant beyond the .01 level. The fact that the latter is so high merely indicates the fact that teachers, in choosing children as 'exceptionally well-adjusted', tended on the whole to select children who were popular. What is more revealing is that, here again, there is no difference whatever as between the means for the two experimental groups and the S.Ds. are comparable. The fact that emotional adjustment is a factor in sociometric ranking has been shown by Dahlke (1953). But it is of some interest to note that the relation is so marked that, even within the two experimental groups used in this study, representing approximately only the lower 10 per cent. of their population, we found correlations of  $-.443$  (aggressive), and  $-.475$  (withdrawing) between

† It is probable that the arbitrary figure of 16 represents a somewhat greater degree of disturbance than category B in the three surveys carried out on behalf of the Underwood Committee, though less than category A.



sociometric ranking and maladjustment score on the Bristol Guides. Both of these are significant at the .02 level. However, although it might be inferred that the two patterns are equally severe insofar as they affect children's relations with one another, it must be admitted that a more detailed study of this aspect of behaviour should be more revealing. Are any of these children actively rejected? How far are their own choices reciprocated? To what extent do they fit in with the 'in-groups' of their respective classes? These are questions which must be left to further investigation.\* If these findings are compared with the ratings given to these same children on the trait of sociability (see Table 2), we see that now there is a marked difference between the two groups, with the aggressive more dependent on company than even the well-adjusted. In some cases this might lead them to feel frustrated by the lack of approval given them by the class as a whole. More often, one may suppose, they tend to form into delinquent or sub-delinquent groups. On the other hand, the tendency to be somewhat unsociable or to prefer being alone which appears in the withdrawing group may be presumed to represent their own response to anxiety—by avoidance.

#### VI.—SOCIAL ADAPTATION.

Our final index of difficulty of adjustment was the Vineland-Manchester Scale of Social Adaptation, an experimental unstandardised revision of the Vineland Scale of Social Maturity.† Because the scale is unstandardised comparisons are internal to the three groups in this study. Absolute figures have no meaning, but the relation between the three groups is of interest.

TABLE 5  
SOCIAL ADAPTATION.

Mean	W	WA	A
Mean .....	53.19	66.33	61.54
N .....	36	36	36
S.D. ....	21.9	19.6	22.5

Previous research has shown that delinquents tend to have low scores on the Doll scale (Dunsden, 1947). It was anticipated that this was largely due to the high weighting of cognitive items in the scale and that the effect would disappear when these items were separated out. This expectation was not borne out, and the relative standing of the three groups remains the same when scores on social perspective items are deducted from the totals. On the other hand the further expectation that the withdrawing group would be particularly

\* W. A. L. Blyth (1959) has shown that it is possible to predict choices in approximately 40 per cent. of cases, using only ecological data. It seems wholly probable that prediction of such choices could be raised to a very high level indeed by combining temperamental criteria with ecological.

† The form used was a try-out for the Vineland-Manchester Scale of Social Adaptation which is being standardised for use with British school children. The standardisation is based on individual interview with the child himself acting as informant. The age coverage is less than the Doll-Vineland scale, 6-15, and within this age range the number of items has been increased. The standardisation sample consists of over 400 children in the North-Western area. Unlike the original scale, the Manchester revision will yield separate indices for various aspects of social adaptation, and in particular, for 'Social Perspective,' which is largely a cognitive attribute, and 'Social Responsibility,' which is practical or executive.

low in social independence is fully corroborated. Results on this scale were compared when the three groups were sub-divided into three age sub-groups, and tested by analysis of variance. The trends were consistent for all ages and the differences proved significant well beyond the .01 level. The conclusion might well be drawn that if some of these children were encouraged to gain greater independence, their adjustment might improve.

#### VII.—ONE-YEAR FOLLOW-UP.

The evidence so far adduced suggests that there are probably more cases of maladjustment which take a predominantly aggressive form, but that the severity of the handicap may be comparable for the two groups as a whole, although on some counts at any rate the valuation of Wickman's clinicians is borne out. However, for the most part the results are less decisive, and the problems as a whole seemed to us sufficiently important to warrant a more decisive form of analysis. This can only be done by means of a long-term follow-up study. In the final analysis those forms of behaviour which are overcome in the course of maturation and experience may be regarded as benign, and those which are not merit our more serious attention. As a preliminary to such an investigation, the teachers concerned were asked to complete a second Bristol Guide in respect of the majority of the children rated as severely disturbed (i.e., those who had been awarded a score of 16 or more in our system of scoring the Guides) and at the same time to grade them on a five-point scale as considerably improved, improved, much the same, somewhat deteriorated, or considerably deteriorated, adding any remarks which they thought relevant. On the basis of their returns it was possible to grade eleven of the withdrawing cases and thirteen of the aggressive on a seven-point scale, as shown in Table 6.

TABLE 6  
FOLLOW-UP OF RESULTS FOR TWENTY-FOUR CHILDREN.

	Marked improvement	Improvement	Slight improvement	No change	Slight deterioration	Deterioration	Marked deterioration
W	—	5	1	4	1	—	—
A	3	3	2	4	1	—	—

It is of some interest to note that if the patterns of behaviour shown in these second returns are compared with the patterns shown one year previously, we find some admixture of the anxiety-hostility pattern in two out of the eleven withdrawing cases, although the predominance of items remains unchanged. Similarly, two out of the thirteen aggressive cases now show some signs of withdrawal or 'unforthcomingness.' In all but one of the remaining cases the pattern remains the same even though the number and severity of symptoms may be reduced. In only one case is there a virtual reversal of pattern. These trends tend to confirm the theoretical presupposition underlying the initial selection. First indications would be that spontaneous improvement occurs in something over half of all cases, in both groups, but that the change may be more spectacular in the aggressive, especially those who, though markedly disturbed, are not seriously maladjusted. However, these conclusions are very tentative and it is hoped to focus our attention on this aspect in future investigations.

## VIII.—REFERENCES.

- <sup>1</sup> BLYTH, W. A. L. (1959). School groups and neighbourhood groups : a study in predictive sociometry. Paper read to Section II (1) c (Education), *Fourth World Congress of Sociology*, Stresa, September, 1959.
- <sup>2</sup> DAHLKE, H. O. (1953). Determinants of social relations among children. *Sociometry*, 16, 327-338.
- <sup>3</sup> DOLL, E. A. (1953). *The Measurement of Social Competence*. Vineland, N.J.
- <sup>4</sup> DUNSDEN, M. I. (1947). Notes on the intellectual and social capacities of a group of young delinquents. *British J. of Psychol.*, 38, 62-66.
- <sup>5</sup> FLEMING, C. M. (1954). *Cumulative Records, Notes on their Content and Use*. University of London Press.
- <sup>6</sup> HOLLINS, T. H. B. (1955). *Teachers' Attitudes to Children's Behaviour*. Unpublished M.Ed. Thesis, Manchester University Library.
- <sup>7</sup> LUNZER, E. A. (in preparation). Aggression and withdrawal in school children ; II—Disparity in attainment.
- <sup>8</sup> MINISTRY OF EDUCATION (1955). *Report of the Committee on Maladjusted Children*. London : H.M.S.O.
- <sup>9</sup> NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH (1952). *A Survey of Rewards and Punishments in Schools*. N.F.E.R. Publications No. 3. London : Newnes.
- <sup>10</sup> PETERSON, D. R., BECKER, W. C., HELLMER, L. A., SHOEMAKER, D. J., and QUAY, H. C. (1959). Parental attitudes and child adjustment. *Child Development*, 30, 119-30.
- <sup>11</sup> SCHONELL, F. J. (1944). *Backwardness in the Basic Subjects*. Edinburgh : Oliver and Boyd.
- <sup>12</sup> STODDILL, R. M. (1931). Parental attitudes and mental hygiene standards. *Mental Hygiene*, 15, 813.
- <sup>13</sup> STOTT, D. H. (1958). *The Social Adjustment of Children*. University of London Press.
- <sup>14</sup> STOTT, D. H. (1959). The prediction of delinquency from non-delinquent behaviour. (Paper read to the Annual Conference of the B.P.S.). *Bulletin of the B.P.S.*, 38, 36A.
- <sup>15</sup> WICKMAN, E. K. (1928). *Children's Behaviour and Teachers' Attitudes*. New York : Teachers' College Contributions to Education.



# SYMPOSIUM: THE DEVELOPMENT OF MORAL VALUES IN CHILDREN

## VII—THE CONTRIBUTION OF LEARNING THEORY.

BY H. J. EYSENCK

(*Institute of Psychiatry, University of London.*)

**SUMMARY.** It is argued that moral values are learnt in the course of the child's development, and that consequently any theory regarding their development should be based on the known facts and principles of modern learning theory. An attempt has been made to do so by suggesting that 'conscience' is, in fact, a conditioned response built up during the child's formative years by the pairing of conditioned stimuli (arising from aggressive, predatory, and overtly sexual actions) and unconditioned stimuli (slaps, beatings, 'shaming', and other punishments) immediately following the conditioned stimulus. Aided by a process of *stimulus generalization* this should, in course of time, lead to an association between the conditioned stimulus and the fear-anxiety responses appropriate to the unconditioned stimulus. Certain deductions are made from this hypothesis and are shown to be supported by experimental evidence. It is argued in particular that if the hypothesis is correct, then individual differences in conditionability, i.e., the ease with which conditioned responses are formed, should be related to moral behaviour and the ease with which guilt feelings are aroused. Here also evidence is quoted to suggest that this hypothesis has some foundation in fact. It is concluded that the treatment which children, adolescents, and adults receive in response to immoral acts, should be based upon the recognition of individual differences between them, particularly with respect to their conditionability, as only in this way will the most beneficial results of such treatment be realized.

### I.—PRELIMINARY CONSIDERATIONS.

THERE is little doubt that attitudes, interests and values are acquired through some process of learning or conditioning during the course of the individual's development, and it seems almost certain from the great deal of evidence accumulated by psychologists, and reviewed by earlier contributors to this symposium, that much of this process of learning takes place during childhood and possibly adolescence. If it is agreed that responses indicative of moral values, whether verbal or behavioural, are indeed, learned responses, then it would seem that modern learning theory, which is probably the most advanced part of psychology, should have a contribution to make to our understanding of their development. In what follows I have tried briefly to indicate the direction which such a contribution might take, and also to quote some experimental data which seem to support the position taken here.

There is an obvious dualism involved in talking about 'moral values'. We may be concerned with the *knowledge* of existing values in a society, and this can be verbally expressed in a questionnaire or an interview; this knowledge is not necessarily related to conduct although there does appear to be a slight tendency for such correlations to exist. Nevertheless, the delinquent child as well as the criminal adult is usually only too well aware of the *fact* that his conduct is contrary to moral precept; his evil-doing is not by and large due to ignorance.

The alternative method of defining the moral values of a person deals with his conduct, rather than with his knowledge, and it would seem that this approach is probably more fundamental and more fruitful than the other. Here also, however, we have a choice to make. We can deal with the matter at the observational and naturalistic level, i.e., by studying actual delinquencies as has been done for instance by Ackerson (1942), or by Hewitt and Jenkins (1946). Alternatively, we can make use of the experimental method, and study delinquent behaviour in the laboratory, in the classroom, or in strictly controlled conditions, as was done for instance in the famous Character Education Enquiry. In either case we discover that among children, adolescents and adults, there appears to be a range of delinquent behaviour from the person who never commits a delinquent or dishonest act, to the person who almost invariably does so. The task of the learning theorist is to account for this gradation in behaviour along the well established lines of learning theory.

Many people interested in the social consequences of crime put the fundamental question in the following way. They ask: "How is it that some people commit crimes when they know perfectly well that their acts are anti-social, and when they have also been taught that crime does not pay?" Such a way of looking at the problem appears to be the matter of putting the cart before the horse. Delinquent behaviour, i.e., the tendency to act out immediately and without restraint one's instinctual impulses, whether sexual, aggressive or predatory, is surely the natural way to act for animals and for young children; the question is: "Why and how do human beings learn to act in conformity with the dictates of society, however inconvenient and distressing these dictates may be, and however much they may conflict with the individual's biological needs and drives?"

It is often suggested that people refrain from immoral acts because of the fear of punishment. It is unlikely that this hypothesis can be taken very seriously. We know from learning theory that the effectiveness of reward and punishment is an *inverse function* of the time interval between act and reinforcement, and a *direct function* of the proportion of reinforcements. Now the one thing we do know about social punishment is that on both counts it fails to provide the necessary mechanism. Punishment usually occurs a long time after the event, sometimes years after, and it only occurs in a small proportion of cases. It is difficult, if not impossible, to give accurate figures about the relative numbers of crimes discovered and punished, as compared with those where the culprit goes free, but the ratio of reinforcement is almost certainly very low.

There are other reasons for doubting the effectiveness of punishment. The well-known and well substantiated Yerkes-Dodson Law tells us that while an increase in drive usually leads to an improvement in performance and learning, there is an optimal point beyond which an increase in drive leads to a *decrement* in performance and learning. The law also states that the more complex the task, the lower is the drive level at which this reversal occurs. Now the kind of punishment meted out by society is usually rather harsh and, therefore, has a high drive level, whereas the task to be learnt, i.e., behaving in conformity with the precepts of society, is obviously a very complex one. Conditions are, therefore, highly unfavourable for punishment to exert its desired influence.

## II.—THE BASIS OF 'CONSCIENCE.'

These and many other reasons make it unlikely that the threat of punishment is the only or even the main reason for moral behaviour. The alternative

suggested by many writers has been a kind of interiorised policeman variously named conscience, 'inner light,' or *super-ego*.<sup>\*</sup> This is usually conceived as some kind of *deus ex machina* implanted in the human being in some mysterious way, which ceaselessly keeps an eye on his activities, and gives him a sharp tweak, whenever he deviates from the straight and narrow path of duty. Descriptively, this is probably not entirely an inaccurate account; the difficulty with it is that there is no known mechanism by means of which such an inner policeman could be called into being, and that the description does not give us any clues about the reasons why some people have a strong and tender conscience while others seem to be completely lacking in it. It is sometimes suggested that environment and teaching are responsible for the apparent individual differences, or that there might be an intellectual deficit which makes some people fail to respond to moral teaching. There is, of course, no doubt that environmental pressures play an important part in the growth of moral ideas and conduct, but such a hypothesis would not account for the frequently observed fact that even the best environment often produces psychopathic individuals apparently lacking completely in any 'inner guiding light,' while the very poorest environment does not by any means invariably produce criminals. Similarly, there is not very much relationship between lack of intelligence and criminality; what relation there is is more likely between low intelligence and likelihood of being found out!

The suggestion made here is a relatively simple one, namely, that *conscience is a conditioned anxiety response to certain types of situations and actions*. In the typical Pavlovian experiment, the dog, through simple pairing of conditioned stimulus (bell) and unconditioned stimulus (meat powder) learns to salivate to the bell, whereas previously it only salivated to the meat powder. Everyone is familiar with this experimental paradigm, and with the fact that similar conditioned responses, particularly of the autonomic nervous system, can be quite easily produced in human beings as well. There is ample evidence to show that anxiety is a conditioned fear response attached to a previously neutral stimulus. Watson's famous experiment with little Albert is probably too well known to need extensive retelling; he induced a phobia for furry animals in an 11-month old infant 'Albert,' who previously had been fond of such animals, by banging a metal bar with a hammer behind Albert's head whenever the infant reached out to pat a white rat (Watson and Raynor, 1920). Learning theory has formalised the rules according to which this conditioning takes place, but we need not be concerned with anything but the bare fact that *anxiety and fear responses can be conditioned in human beings with very great ease*. It is also to be noted that such conditioned responses, once they are formally established, do not extinguish by themselves in the course of time, but require an experimental process of *extinction* (Osgood, 1953.).

<sup>\*</sup> The Freudian conception is concisely given in Freud's "General Introduction": "It is not to be doubted for a moment that one may recognize in the Oedipus-complex one of the most important sources for the consciousness of guilt with which neurotics are so often harrassed. . . . Perhaps mankind as a whole has, at the beginning of its history, come by its consciousness of guilt, the final source of religion and morality, through the Oedipus complex." Perhaps. Even if the first part of this quotation had some factual reference, it would still not be clear how a phenomenon assumed to be of universal occurrence, like the Oedipus complex, could be used to account for individual differences in proneness to guilt feelings. Nor is it at all clear how such a theory could be experimentally or even observationally tested. Possibly one might deduce that boys brought up by widowers whose wives died in childbirth should all develop into psychopaths. At least they should stand in little danger of developing neurotic guilt feelings!



The application of this well-documented process to the development of moral behaviour can be briefly indicated in the following way. A young child behaves in a socially undesirable manner, i.e., by being aggressive, by indulging in overt sexual activity, by stealing, lying and cheating, or in whatever way anti-social behaviour is defined in a given society. There is an immediate sharp punishment—a slap, withdrawal of some privilege, 'shaming', exclusion from the family circle or whatever it may be. This punishment produces pain and fear and the associated autonomic disturbances; these in turn become attached to the *type of situation* and the *type of action* which called forth the punishment, thus producing a conditioned anxiety reaction whenever similar situations and actions re-occur.

It might be argued that surely identical situations never occur, but this objection is taken care of by the well-known fact of *stimulus generalisation*. Little Albert was taught to be afraid of a rat, but this conditioned fear generalised to rabbits and other furry animals. Stimulus generalisation, therefore, will account for the fact that not only identical but similar situations and actions will also call forth the conditioned anxiety response in the child which has been punished once or several times for certain types of misdemeanour. This process of generalisation is undoubtedly aided by the fact that parents often draw attention to similarities between different antisocial acts by a process of *labelling*. There is ample experimental evidence to show that generalisation of conditioned responses does proceed along verbal lines. Thus the subject who has formed a conditioned P.G.R. response to the word 'cow', because this word was in the past followed by an electric shock, will now have a conditioned P.G.R. response also to the names of other animals, such as 'goat'; he will not, however, generalise responses to words such as 'how' which are, apparently, much more similar to the original conditioned stimulus from the point of view of sound.

We thus have a growing child in whom conditioned anxiety responses have been built up to anti-social behaviour in situations involving aggression, sex, etc. When temptation arises, there also arises the conditioned anxiety, and we, consequently, have some form of hedonic calculus involving not *present* satisfaction and *future* punishment (in which present satisfaction would almost certainly win), but rather *present* satisfaction in opposition to *present* discomfort as produced by the anxiety reaction in response to overt anti-social behaviour. If the conditioning has been strong enough, 'conscience' will win the day, and the individual will withdraw from the situations without giving way to 'temptation.' Even if he does give way to 'temptation' there will still be a strong anxiety reaction to detract from his enjoyment, thus making it less likely for him to react anti-socially on the next occasion.

It might be queried whether anxiety is in effect strong enough for this proposed role. When we consider that many people have committed suicide rather than suffer strong anxiety, and when we consider that criminals have, on many occasions, preferred to give themselves up and take their punishment, rather than continue to bear the anxiety produced by their crime, we may not feel so doubtful about the efficacy of the proposed mechanism or its strength. However, such facts, of course, do not provide proof of the hypothesis in question; they merely illustrate the strength which feelings of anxiety can reach. We must now turn to the evidence firmly supporting our hypothesis.

### III.—INDIVIDUAL DIFFERENCES.

It is well-known that there is a group of people, not sharply segregated from the rest, but presumably continuous with the remainder of humanity, which is characterised by anti-social behaviour in almost pure culture. I am

referring, of course, to the so-called *psychopaths*, i.e., to children, adolescents and adults, who, in spite of often high intelligence and good up-bringing, seem to be completely lacking in moral sense—so much so that they have even been called *moral imbeciles*. They will lie, cheat, steal, rape, and indulge in any form of anti-social activity without apparent regard for consequence, and without regard for their victims. Such people will often commit criminal acts for the slightest of gains, and in situations where discovery and severe punishment are practically certain. Here is a group of people for whose behaviour no adequate theory has been put forward. Can our theory do better?

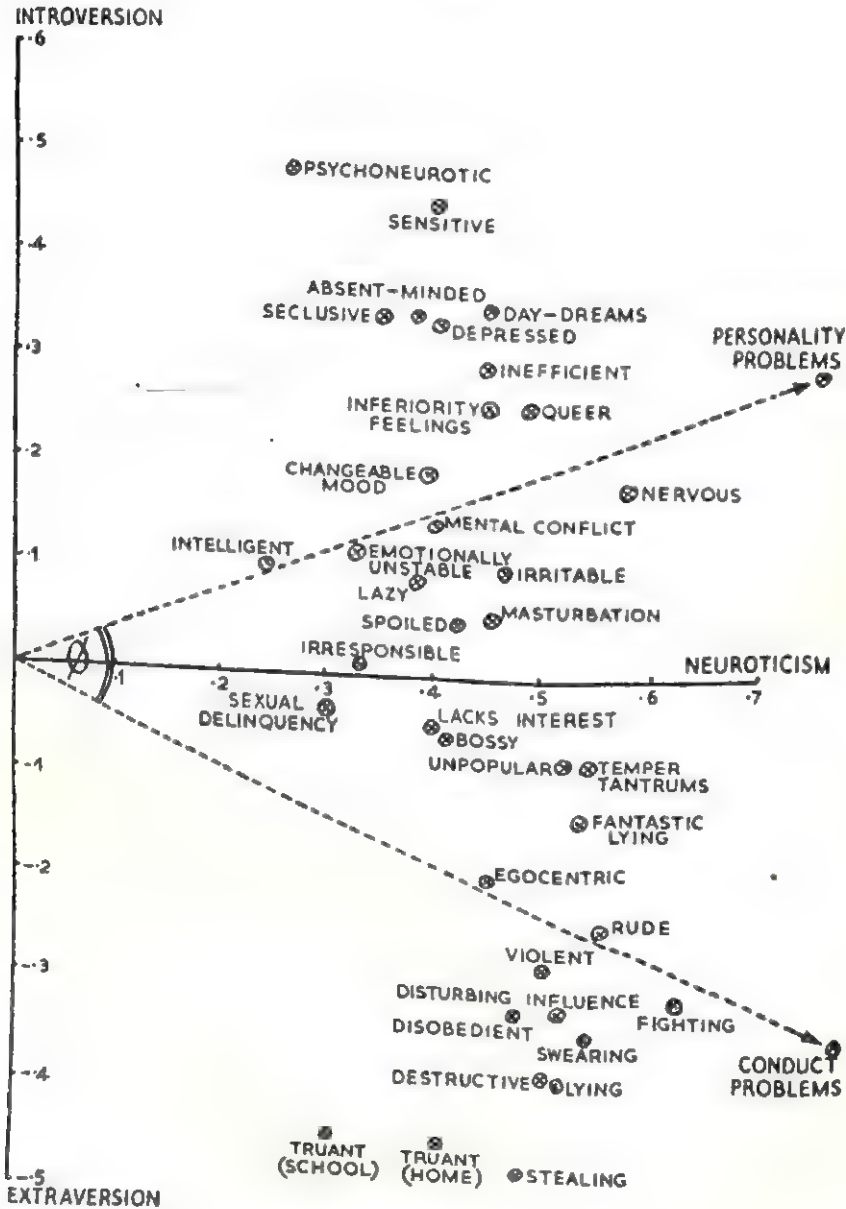
Let us note, first of all, that Pavlov already discovered very great individual differences in respect of *conditionability* between his dogs. Some dogs are very easy to condition, others very difficult, with the remainder in-between. This finding has been universally verified in the case of other animals, and also in the case of human beings. Differences in conditionability are by and large unrelated to such factors as age, sex and intelligence. If our theory is accurate, then we might be able to explain the behaviour of psychopaths by postulating that such people are endowed with a nervous system which is largely resistant to the formation of conditioned responses. If this were so, then the course of up-bringing would not suffice to produce in them the requisite anxiety responses, which we have postulated to lie as the basis of 'conscience.' It should be easy to test this hypothesis by attempting to condition psychopaths in the laboratory, and by comparing their success or failure with the responses made by an average group of people put in the same experimental situation. When this is done, the results very strongly bear out the hypothesis, and it is found that compared with a normal group of people, psychopaths are, indeed, very difficult to condition, requiring many more pairings of conditioned and unconditioned stimulus before any effect is seen (Lykken 1957, Eysenck, 1957). Results, therefore, are in accordance with our hypothesis.

We might put forward another hypothesis to complement the previous one. There is a group of people in society who are suffering from unduly strong anxieties related to stimuli which, in the normal person, do not arouse anxieties at all. These people suffer from fears of open spaces, fears of enclosed spaces, fears of animals, and so forth, without being able to give any rational account of their fears. Such people constitute the main group of neurotics, whether in-patients or out-patients, at our hospitals and clinics; they are sometimes known as 'anxiety states.' Such people also frequently show strong guilt feelings about actions which few people would regard as immoral or anti-social in any real sense of those terms. In other words, these people appear to have a conscience much more tender than the average person. In terms of our hypothesis we would expect such people to be particularly easy to condition; their symptoms and their behaviour could then be accounted for in terms of a too ready conditioning of fear responses to a large number of previously neutral conditioned stimuli (Cf. fn. p. 000).

Again, the evidence strongly supports this view. The work of Spence (1956) Franks (1956, 1957), and many others (cf. Eysenck, 1957) shows that people suffering from anxiety states are more easily conditionable than the average run of people, and accordingly, we have a complementary piece of evidence showing that we can arrange human beings in a continuum of 'conditionability' from high to low, a continuum which runs parallel with a continuum of behaviour patterns going from hyper-moral through average to psychopathic and immoral. There is more evidence in favour of this hypothesis, and some of it has been quoted elsewhere (Eysenck, 1957). Instead of going into this additional evidence I would prefer to draw attention to another extension of this general scheme.

I have argued that the personality dimension, at the one extreme of which we have the psychopath, and at the other extreme of which we have the anxiety state, can be found not only in the emotional, anxious and neurotic type of personality, but also in the population as a whole, and I have suggested that we are, in fact, here dealing with the well-known extravert-introvert typology transferred into the field of neurosis and maladaptation. Figure 1 will illustrate

FIG. 1





this notion ; it is quoted from " The Structure of Human Personality " (Eysenck 1960) and it is a diagrammatical presentation of a factor analysis of fifty traits which had been correlated by Ackerson (1942) among several thousand children who had been studied at the Illinois Institution for Juvenile Research. It will be seen that of the items included, all have correlations with the general factor of neuroticism or emotional instability, but that there is a division along the lines of the second factor into *personality problems* characteristic of the introvert, and *conduct problems* characteristic of the extravert.

Much the same kind of result was found by Hewitt and Jenkins (1946) in a study of 500 problem children. When they correlated various notations from the case histories of these children, they found three main characteristics, ranging from 'unsocialized aggression' (the extraverted end) through 'socialized delinquency' to 'over-inhibited behaviour' (the introverted end). Other data are quoted in " The Structure of Human Personality," to support the existence of a general dimension of this nature. There is no direct evidence of a relationship between these behaviour patterns in children and their conditionability ; here would seem to be an area of research which could, with advantage, be pursued by those responsible for the moral welfare of children and for their disposal after they have come in contact with the law.\*

Among adults, however, there is considerable evidence in the work of Franks (1956, 1957), and others (Eysenck, 1957) to show that there does exist a significant relationship between introversion and conditionability, even when extreme cases of psychopathy are not included in the neurotic sample, and even when quite ordinary normal subjects are being tested. In so far as our hypothesis deals with conduct and patterns of observable behaviour it must, I think, be concluded that we are on relatively safe ground in putting forward the hypothesis that *differences in conditionability determine in part the socialized or anti-social behaviour of children and adults alike.*

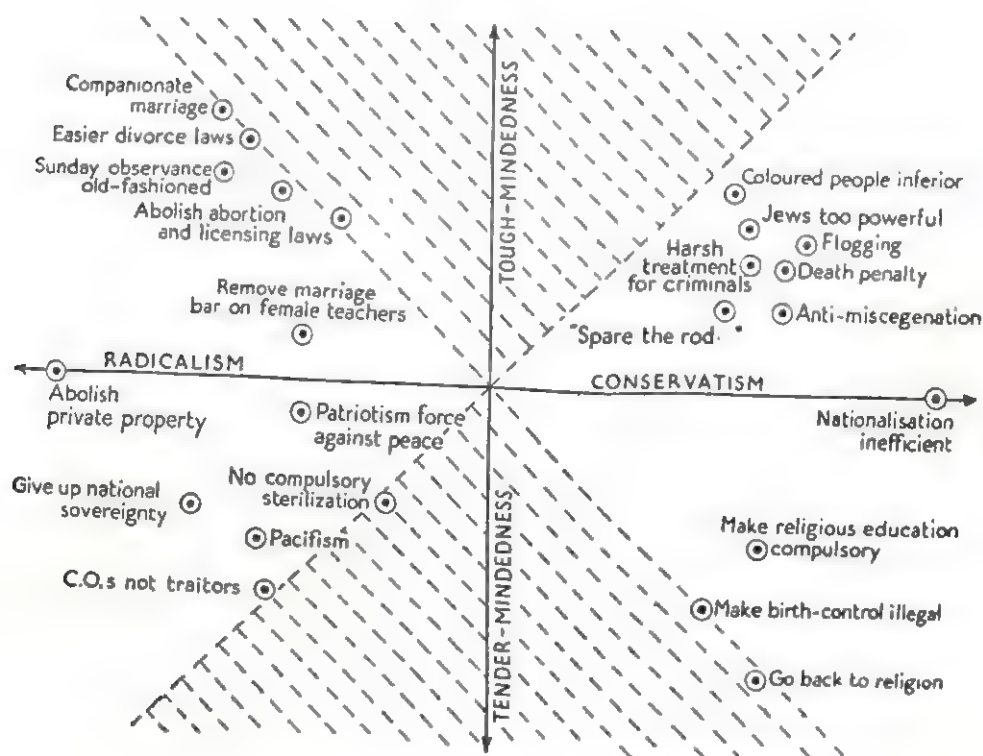
#### IV.—IMPLICATIONS OF THE THEORY.

Does the theory have anything to say about the more verbalized type of values and attitudes ? An attempt has been made in *The Psychology of Politics* (Eysenck, 1954) to make certain deductions from the theory, and to test these objectively. It would seem to follow from the theory developed here that introverts would be more concerned with ethical and moral prohibitions, with religious ideas and quite generally with the erection of barriers against the direct and immediate satisfaction of instinctual and libidinal impulses. Conversely, it would seem to follow that extraverts would be more likely to favour the direct expression of sexual, aggressive, and other anti-social impulses and to be less concerned with ethical, moral and religious ideas. The proof of this

\* The reader may wonder why both personality problems and conduct problems have such high loadings on neuroticism, and whether adult criminals too would be found to be neurotic. It would appear from some recent work done with the Maudsley Personality Inventory on recidivist criminals, that their neuroticism is indeed almost as high as that of hospitalized neurotics. Some learning theorists, particularly Spence, regard neuroticism or emotionality as a kind of drive variable (D), which according to Hull's theory, is multiplied by habit ( $s^E_R$ ) to produce behaviour ( $s^E_R$ ). If we regard, in a very rough-and-ready fashion, the extravert as a person in whom 'temptation' > 'conscience', whereas in the introvert 'temptation' < 'conscience', then the added drive produced by the emotionality-neuroticism variable might be thought to produce a much stronger reaction in the direction determined by the respective weight of these two influences. This is a highly speculative consideration and it would require a considerable amount of experimental work to decide if any real meaning is attached to it.

hypothesis requires two series of studies. In the first series, a number of correlation studies were carried out on the attitudes of quite large numbers of subjects towards a number of variegated social issues. When these attitudes were inter-correlated, they invariably gave rise to two main factors, the ever present one of radicalism-conservatism, which does not concern us here, and a second factor, independent of the first, the two poles of which were entitled tough-mindedness and tender-mindedness. The factor structure is illustrated in Figure 2, and it will be seen that tender-mindedness is a factor which appears to be characterised by those attitudes theoretically described as introverted, while tough-mindedness is made up of attitudes theoretically attributable to the extravert. Thus, a tough-minded person favours the overt expression of aggression (by flogging, death penalty, birching, etc.), towards out-groups (criminals, Jews, coloured people, etc.), and the overt indulgence in sexual activities (companionate marriage, easier divorce laws, the abolition of abortion laws). The tender-minded person has strongly favourable ideas towards religion and ethical ideals such as pacifism, etc. The first step in our proof, therefore, does not seem to contradict the hypothesis (Eysenck, 1954).

FIG. 2



As the second step, several studies have been carried out, as I have mentioned in the *Psychology of Politics*, showing that correlations do indeed exist between tender-mindedness and introversion on the one hand, and tough-mindedness and extraversion on the other. Some of these studies have been criticised because of the non-representative nature of the samples, and consequently, I have in a recent investigation, repeated this work on a

random sample of the population (Eysenck, 1960). The results were fully in line with those previously reported and seem to leave little doubt about the existence of some such relationship as that postulated. It would appear, therefore, that in the field of verbally expressed moral attitudes and ideas also the general hypothesis here put forward can make possible verifiable predictions.

It will, of course, be obvious to the reader that there is no intention, in sketching out this theory, to account for all criminal or immoral behaviour. The nature of such behaviour, and the manifold determining causes which are involved are too complex for any single factor theory to be acceptable to students of the field. It is obvious that individual differences in conditionability are important in deciding whether a particular conditioned response will or will not be established; it is equally obvious that for such a response to be established, even in the most easily conditioned person, there must first of all occur a *process of conditioning*, i.e., of pairing the conditioned and the unconditioned stimulus. Thus, in conditioning we are obviously involved in what one vaguely calls the *social factors* of character building, and it is clear that these differ from person to person, and probably quite generally from class to class. Thus, there is some evidence from American researches, particularly the work of Kinsey (1948) and the Chicago school, that lower-class groups lay far less emphasis on the extinction of overt aggressive and sexual reactions than do middle-class groups. In these circumstances we would expect quite a different pattern of behaviour in middle and working class boys, even if there were no differences in conditionability between them, and these differences are, indeed, found. (Cf. also Eysenck, 1951.)

Another point where environmental influences are clearly of the greatest importance is that relating to the strength of a given temptation which has to be overcome by the conditioned anxiety response we call 'conscience.' If, indeed, a person's reactions are determined by the respective strength of 'conscience' and 'temptation,' then clearly the same person may act morally on one occasion and immorally on another, depending on the degree of desire, hunger, anger, or whatever may be involved in the situation. A poor man, as has often been pointed out, is more likely to steal a loaf of bread than a rich man, because the temptation for him is so very much stronger; from comparing the actions of the rich and the poor in respect of stealing a loaf of bread, obviously no deductions can be made as to their likelihood to form conditioned responses easily. All these considerations will be obvious, but I have mentioned them briefly to avoid criticisms sometimes made of theories of this kind, to the effect that they do not take into account all the conditions of a complex social phenomenon.

Nevertheless, our theory demands that at least to some extent we should recognize, and be prepared to discover, *innate biological factors* determining in some degree the moral or immoral, criminal or non-criminal reactions of human beings to certain types of situations. Conditionability, presumably, is a function of certain features of the central nervous system, and as such is likely to owe much to heredity. Is there any evidence about the hereditary determination of crime? The important and, indeed, fundamental work of Lange (1931), Stumpfl (1936), and Kranz (1936) has shown that such hereditary determination can, indeed, be demonstrated very clearly. As is well-known, these investigators located prisoners who had a like-sexed fraternal or identical twin. They then investigated this twin to see whether or not he also had been convicted of a crime, and if so, whether his crime was similar to that of his sibling. They found a very considerable degree of concordance for identical twins, and



much less for fraternal twins; it is difficult to interpret the evidence in any other way, but as a strong confirmation of the heredity hypothesis. This conclusion is well expressed in the title which Lange gave to his book: *Crime as Destiny*. It may be noted that J. B. S. Haldane wrote an introduction to this book in its English translation, in which he fully concurred with the author's conclusions; this is relevant and important because of his expert knowledge in the field in genetics and because of his well-known political sympathies which would make him less likely to accept such conclusions with enthusiasm.

Do any practical suggestions follow from the theory here developed? Two points seem to be worth making. In the first place, there has been far too little work of an experimental nature directly related to this problem to make it possible to decide with any degree of conviction as to the adequacy of the theory here presented. What is needed is a decade of concentrated experimentation before we shall be able to come to a conclusion with any degree of confidence. It would, therefore, appear to be premature to make any detailed practical suggestions; these must await further qualification of theoretical issues and further experimental verification.

In the second place, however, it does seem to me that if there is any degree of truth at all in the theory here presented, then it does lend some support to the slogan that punishment should not fit the crime but the criminal. Perhaps we might amplify this a little and say that character education cannot and should not be a uniform process, but that it should take into account the individual personalities involved. This saying, of course, is a truism which has often been repeated; what I am suggesting is that it can only cease to be a truism, and become a guide to action, when we know what are the parameters directly involved in moral actions, and how these parameters can be measured, and if possible, affected by our actions. If it were true that conditionability plays as important a role as I have suggested, then it would seem to follow that the maxim "spare the rod, spoil the child" could with advantage, be applied to the extraverted, possibly psychopathic, non-conditioner, whereas the modern free-and-easy methods of up-bringing would be much more appropriate to the introverted, anxious, easy-to-condition type of child. It may be possible that the acceptance of some such general rule might reduce the number of both behaviour problems and personal problems, which at the moment, appear to be increasing to such an alarming extent. However, on this point also much further research, possibly of an applied nature and carried out in the classroom, will be required before we can say for certain whether this suggestion is likely to have the consequences envisaged.

## V.—REFERENCES.

- ACKERSON, J. (1942). *Children's Behaviour Problems*. Chicago: Univ. of Chicago Press.
- EYSENCK, H. J. (1951). Primary social attitudes as related to social class and political party. *Brit. J. Sociol.*, 2, 198-209.
- EYSENCK, H. J. (1954). *The Psychology of Politics*. London: Routledge and Kegan Paul.
- EYSENCK, H. J. (1957). *The Dynamics of Anxiety and Hysteria*. London: Routledge and Kegan Paul.
- EYSENCK, H. J. (1960). *The Structure of Human Personality* (2nd ed.). London: Methuen.
- EYSENCK, H. J. (1960). Personality and social attitudes. To appear. *J. Soc. Psychol.*
- FRANKS, C. M. (1956). Conditioning and personality: a study of normal and neurotic subjects. *J. Abnorm. Soc. Psychol.*, 52, 143-150.
- FRANKS, C. M. (1956). Recidivism, psychopathy and personality. *Brit. J. Delinquency* 6, 192-201.
- FRANKS, C. M. (1957). Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 48, 119-126.

- HEWITT, L. E., and JENKINS, R. L. (1946). *Fundamental Patterns of Maladjustment*. Illinois: D. H. Green.
- KINSEY, A. C., POMEROY, W. B., and MARTIN, C. E. (1948). *Sexual Behaviour in the Human Male*. London: Saunders.
- KRANZ, H. (1936). *Lebensschicksale Krimineller Zwillinge*. Berlin: Springer.
- LANGE, J. (1931). *Crime as Destiny*. London: Allen and Unwin.
- LYKKEN, D. T. (1957). A study of anxiety in the sociopathic personality. *J. Abnorm. Soc. Psychol.*, 55, 6-10.
- OSGOOD, C. E. (1953). *Method and Theory in Experimental Psychology*. Oxford: Univ. Press.
- SPENCE, K. W. (1956). *Behaviour Theory and Conditioning*. New Haven: Yale Univ. Press.
- STUMPFL, F. (1936). *Die Ursprünge des Verbrechens, dargestellt am Lebenslauf von Zwillingen*. Leipzig: Thieme.
- WATSON, J. B., and RAYNOR, P. (1920). Conditioned emotional reactions. *J. Exp. Psychol.*, 3, 1-14.

(Manuscript received, March 20th, 1959.)



# GOAL-SETTING BEHAVIOUR, ANXIETY AND SCHOOL STREAMING

By THELMA VENESS  
(*Birkbeck College, London*)

**SUMMARY.** A particular pattern of goal-setting behaviour in a level of aspiration test was found to occur more frequently in technical schools and in the "A" streams of modern schools than in the lower modern school streams and in grammar schools. In previous research, this pattern, of setting high goals and adhering to them rigidly, whatever the fluctuations in performance, has been associated with anxiety and similar personality traits. Matched pairs of these children and controls were compared for other possible indices of anxiety in the material of an investigation of which this is a part. The findings are discussed in relation to the possible effects of success and failure involved in secondary school selection.

## I.—INTRODUCTION.

THE work to be described forms a small part of an enquiry into the aims and aspirations of school-leavers; their vocational and other ambitions; their ideas, values and plans concerning the future. In this enquiry, directed by Professor C. A. Mace and assisted financially by a grant from the Nuffield Foundation, several techniques of enquiry were directed at young people at different types of school, all expecting to leave school in about twelve to eighteen months. Among the techniques used was a level of aspiration test. The essential requirement of such a test is a task which can be repeated several times so that the subject, on the basis of his achievement on any trial, can set his goal for the forthcoming one. The conditions of the present investigation demanded a task suitable for administration to whole classes of children, so that each child could, without help, rapidly calculate his performance at the end of the trial and record his aspiration for the following attempt. In addition, previous research<sup>5</sup> has indicated that different goal-setting behaviour may be demonstrated according to whether the task resembles a scholastic exercise or a game; for that reason, a task was selected that was thought to have elements of both, namely, the making of words from the letters of a given long word. For classroom administration, booklets were made up, each page headed by an eight-lettered word, the words having been previously equated for difficulty. The rules of the Word Game were explained and the children given two minutes in which to make up words. At the end of this period, they were asked to count the number of words written (correctness being of no importance in this situation), to record this score at the bottom of the page and then to write the number of words they thought they would make next time. There were eleven such words, yielding ten samples of goal-setting for each child. A particular pattern of goal-setting behaviour is the subject of this paper.

## II.—THE CRITICAL FINDINGS.

From the performance and aspiration scores so obtained, several numerical relationships can be calculated and used as indices for each subject. Of importance for the present enquiry are two:

(a) The goal discrepancy score: the mean of all the differences between aspiration level and the immediately preceding performance score. In using



this index, it is assumed that the subject decides his next goal on the basis of his present score; i.e., that this index represents numerically the subject's liability to aim at targets above or below his likely attainment.

(b) A measure of rigidity of goal-setting, or the tendency to remain at the same aspiration level regardless of fluctuations in actual performance. Here, a child was classified as 'rigid' if not more than two different numbers were used in stating the ten goals. Since serial order was disregarded for this purpose, the following pattern, 10, 9, 9, 10, 10, 9, 10, 9, 9, 9, would qualify for inclusion. In no case so classified was the actual performance so inflexible as to justify the rigidity of goal-setting.

Of particular interest is the combination of high scores in both these measures, a pattern of setting a goal above likely performance and of then adhering closely to that goal throughout the test. In terms of the present data, this pattern represents the concomitance of the rigidity measure already described and a mean goal discrepancy score above the median of the total sample (+1.0). Of the 778 children tested, in schools of one local education authority (a Home County), forty-four children gave this pattern of responses—a proportion of 5.6 per cent. Several had a very high goal discrepancy score—as high as eleven in one case: thirty-eight were in the upper quartile for the whole sample. Eleven children made no change at all in their choice of goal.

This goal-setting pattern, although uncommon in the present sample, merits attention because of its resemblance to those described by other investigators and associated by them with certain personality variables. P. S. Sears,<sup>8</sup> working with younger children, found that such goal-setters had lower ratings on self-confidence, made lower estimates of their performance on various school activities though this was associated with a higher wish for achievement, and on a questionnaire of emotional stability were more likely to indicate anxiety, tension and poor school adjustment. She describes children in this group thus: "In general, they are not very well adjusted socially; are tense, apprehensive, sensitive, insecure, worried and usually seem to utilize persistent non-adjustive forms of reaction to problems of their school and social life."<sup>8</sup> Eysenck and Himmelweit<sup>4</sup> studied a similar pattern in adults and found it to belong largely to 'the affective group,' a term referring to a syndrome, later called 'dysthymia,'<sup>3</sup> the main characteristic of which is anxiety. D. R. Miller<sup>6</sup> described the neurasthenics in his test population as being more likely, when threatened with failure, to predict high scores for themselves, and scores which do not change in relation to the changing situation. 'Neurasthenics' are here defined by vegetative symptoms such as sweating, tremor, diarrhoea—symptoms usually associated with anxiety.

This pattern of goal-setting behaviour would seem, then, to be slightly pathological. The fact that only 5.6 per cent. of the sample show it is a welcome finding. Of particular interest, however, is the further finding that the forty-four children in this group do not represent equally the types of school and school stream tested.

The first difference to be noted is between the proportions of what we might, for brevity, call 'anxious' goal-setters found in the A streams of the seven modern schools and those in the lower streams of the same schools. A higher proportion of such children were found in the A classes, to a degree significant beyond the 1 per cent. level, using Brown's modification of the Chi Square test<sup>2</sup> and correcting for discontinuity. There was no discrimination according to age, sex or social class.

The first explanation of this difference that comes to mind relates it to intelligence, the more intelligent children being readier to adopt high, unattainable goals and maintain them with tenacity. This hypothesis is not supported, however, when the grammar and technical school results are examined. The proportion is even higher in the technical schools whereas it is low in the grammar schools. Technical schools and modern A streams combined have a higher proportion, significant beyond the 1 per cent. level, than have grammar schools and the rest of the modern streams combined. The relevant proportions are shown in Table 1.

TABLE 1  
PROPORTIONS OF 'ANXIOUS' SCORERS.

School Stream	N	Anxious Scorers	
		No.	%
Grammar .....	200	8	4.00
Technical .....	125	12	9.60
Modern A .....	201	18	8.95
Modern B, etc. ....	252	6	2.38

In seeking further explanations, and having regard to the possible link with anxiety, one is tempted to speculate about the effects on children of their finding themselves in certain types of school, particularly as their placement results largely from an entrance examination which is well-known to give rise to anxiety in parents, anxiety which might well have been transmitted to the children. This last point is made in the report on Secondary School Selection issued by the British Psychological Society<sup>9</sup>, which refers also to cases, admittedly rare, of breakdown alleged to be associated with the selection examination. 'Success' is very widely measured, by the general public, by the attainment of grammar school standard, so that even the child in the technical school, like the child in the modern school, may feel a failure. It can be argued further that the modern school children of ability lower than that found in the A streams might never have hoped to 'pass' the examination and for that reason had no subjective feelings of failure. The feelings of failure would, therefore, be stronger with the technical school children and those in the A streams of modern schools—the very places where we found the higher proportions of 'anxious' scorers.

It is appreciated that these suggestions do not accord with comments in the book already cited,<sup>9</sup> in which it is argued that despondency among grammar school rejects has been exaggerated; that it is grammar school children who are more liable to be anxious and worried by the demands that the curriculum is making upon them. On the other hand, Sears<sup>7, 8</sup> claimed that among the group of children she tested, those with the pattern of goal-setting under discussion were usually the children she classified as having a past record of academic failure, although in this case, 'failure' referred to the type of task which constituted the level of aspiration measure.

An obvious objection to this hypothesis is that the effects of an examination taken at the age of ten would be unlikely to be active still at the age at which this investigation was conducted—between four and seven years later. Some tangential material on this is to be found in other data collected from the children at the same time. They were asked to write a brief account of what was

entitled 'The Best Moment of my Life.' A considerable proportion (14 per cent. of boys, 13 per cent. of girls) in the grammar school, sometimes seven years after the event, wrote of the day when they had heard of their success in the 'scholarship.' Unfortunately, the complementary evidence is not available, i.e., the opinions of the other children on 'The Worst Moment of my Life'!

These speculations cannot be pressed further at this point: unlike manufacturers of cereals, we would need a great deal of hard evidence concerning the causes and effects of success at the selection examination. Further, it might well be suggested that the evidence we have cited from the level of aspiration test is a very slender foundation for a speculative superstructure attempting to explain anxiety in school children. The first task is to study these forty-four children further to see whether one is warranted in regarding them as anxious. Opportunities for a close examination of the individual children were not available. The material at hand, therefore, i.e., the material collected in the main investigation, had to be scrutinised for further clues, and this scrutiny comprises the remainder of this paper.

### III.—PLAN OF FURTHER INVESTIGATION.

To make more rigorous the search for further signs of anxiety, the forty-four children were compared on various indices with a control group of children matched, child for child, in several respects. The matched pairs came from the same class; this controlled for level of schooling and for sex, and adequately for age and general level of ability. They were matched, also, on certain level of aspiration measures. Thus, every control child had a positive goal discrepancy like its pair. As far as possible a child was chosen with a goal discrepancy score approximately equal to that of the child to be matched, but the 'anxious' group tended to be so high in their goals that a precise match was not possible. The total goal discrepancy score for the critical group was 131 and for the control group, 100. They were further matched on another level of aspiration measure not yet mentioned, which might be called a measure of reality. It is an index derived from the supposition that experience of 'failure' (that is, falling short of the target adopted) should not lead to a raising of one's goals, nor an experience of success to a lowering of one's goals. It is not necessary to describe in detail here how the index was calculated, except to say that it was derived in such a way as to be independent of the measure of rigidity which supplies the crucial difference between the groups. An alternative estimate of this rigidity can be calculated from the total number of shifts in goal made. On this measure, the rigid group has a total of seventy-six shifts and the control group 216.

In the next section are listed seventeen specific items on which the two groups were compared, together with the prediction made and the results found. A methodological point to be noted here is that, for many of these items, the individual child's score had already been assessed, independently of the present investigation. Where this was not the case and a new assessment was needed for testing a specific hypothesis, the two groups were put together so that the investigator did not know which were the 'anxious' children.

Before discussing the specific items, the five techniques of enquiry from which they were drawn must be described. They are as follows:

A.—An essay in which the children were asked to imagine that they were near the end of their lives and tell their life stories from the time they left school. Time checks were given after 30 minutes to ensure that during the last ten minutes they would give attention to the end of the story. Items 1 to 7 refer to this material.



B.—A ten minute essay entitled "The Best Moment of my Life." Items 8 and 9 are taken from this.

C.—A questionnaire in the form of twenty-two statements bearing on values related to jobs. Each statement was first marked as Important or Not Important in one of two columns. The children then marked in two further columns those that were Very Important to them, and then the two that were Most Important. A weighted score for each statement is derived. Items 10 and 11 refer to two of these statements.

D.—A questionnaire of the usual form from which come items 12 and 13.

E.—An individual interview from which items 14 to 17 are taken.

#### IV.—RESULTS.

The seventeen items and the actual finding on each are as follows:

(1) Balken and Masserman,<sup>1</sup> comparing the lengths of Thematic Apperception stories given by neurotics of three kinds, found that those of people classified as anxiety cases were the shortest. It was predicted here that the essays of the 'anxious' children would be shorter. A word count was made, but no significant difference was found using the Wilcoxon Test: in fact, in only twenty-three out of forty-two pairs did the 'anxious' child write a shorter essay. (Occasional absence from school for parts of the investigation is the reason for not all results being out of forty-four.)

(2) A further finding of Balken and Masserman<sup>1</sup> was that the stories of their anxiety group had a higher verb to adjective ratio because, it was said, they were more intense and dramatic. It could be argued, perhaps, that the proliferation of adjectives implies a relaxed, leisurely attitude unlikely to be found in anxious people. Accordingly, a verb to adjective quotient was calculated for the two groups, again from the essays, a higher quotient for the 'anxious' children being predicted. Again, the prediction was not supported by the Wilcoxon Test: nineteen out of forty-two pairs had differences with the sign contrary to hypothesis.

These two departures from the findings of Balken and Masserman cannot, of course, necessarily, be taken to weaken their findings. They were not dealing with written material; their stories were spoken. Different mechanisms may operate in this situation from one in which stories are written down, and where there is a known time limit.

(3) It was predicted that, in the telling of a life story, anxious children would be more likely to refer to the savings made by themselves or their immediate families. This is an *a priori* hypothesis, assuming that anxiety will be associated with seeking security, and that 'saving' is a commonly recognised avenue to financial security. Savings were mentioned by nine of the forty-four 'anxious' children compared with eight controls, a difference not statistically significant.

(4) Mention of earnings in the life story was treated in the same way, with the prediction, on similar grounds, that more frequent references would occur in the 'anxious' group. Again, the difference is not significant, there being nineteen 'anxious' children and eighteen controls out of forty-four.

(5) A further tally from the life stories is the incidence of mention of accidents and disasters. The prediction here was that the 'anxious' children, again guided by their clinging to security and avoidance of situations evoking anxiety, would be less likely to describe such events in their projected life

stories. Here, the results are in the predicted direction—eleven 'anxious' *versus* sixteen controls out of forty-four—but do not reach statistical significance.

(6) The essays had also been scored previously for evidence of conflict such as uncertainty about a choice of career. It was predicted that the incidence would be higher with the 'anxious' group. The figures are seven (the 'anxious' group are always given first) as against four out of forty-four, not a significant result.

(7) The final item from the essays is the incidence of mention of their own parents. The prediction—again on the grounds that security is important to the 'anxious' group—was that the incidence would be higher here. This was so, the figures being sixteen and eleven out of forty-four, but the difference does not achieve significance.

(8) With the suggestion in mind that an examination had been of importance to these 'anxious' children, reference was sought in the "Best Moment of my Life" essay, to examinations of any kind, in the expectation that the incidence would be higher for the 'anxious' group. In fact, the figures were only six and five out of forty-two.

(9) This same short essay had previously been scored for themes concerned with 'relief from anxiety',<sup>10</sup> which would include, for example, recovery from illness or the return of a parent after a separation. It was expected that such themes would occur more frequently in the 'anxious' group. This was so, but the proportions were too small for statistical test, being seven and two cases out of forty-two.

(10) This and the next item relate not directly to anxiety but to obsessionality, and its concomitants. To maintain a goal rigidly in a level of aspiration test might well be regarded as obsessional behaviour, and Eysenck's dysthymic group does contain obsessives as well as anxiety cases. From the questionnaires described above (C), two items were extracted, the items which, in the whole investigation, seemed to have the closest reference to obsessive-compulsive tendencies. On the assumption that the 'anxious' group is in fact anxious rather than compulsive, the prediction was made that it would *not* score higher than the control group on these items.

Item 10 refers to the statement, "I should be able to keep clean and tidy at work." As described above, a weighted score is extracted for these items each child can score from 0 to 3 according to the degree of importance attached to it. The weighted scores total 56 and 59 for the two groups, supporting the prediction.

(11) Here the statement is "I should be able to finish one job before I have to think about the next." The weighted scores of the 'anxious' group total forty-one, of the control group fifty-nine. On a Chi Square test, comparing the distributions of scores, from 0 to 3, the tendency of the 'anxious' group to score lower is confirmed at the 0.5 per cent. level.

(12) Prior to the present investigation, an assessment of ambitiousness had been made on the basis of four questions, asking for expectations of the first full time job, the job after two or three years, after nine or ten years, and just before retirement. Signs of upward social mobility were the criteria of ambitiousness. The prediction made here, suggested chiefly by the high goal discrepancy score of the 'anxious' group, is that more of them would have been scored as ambitious. In fact, fourteen out of forty-four of them were, as against ten of the controls. The difference is in the predicted direction, but not significant.

(13) A rather similar assessment had been made of ambitiousness, but this time from three questions about what the child wanted to own at different stages of his life. Here, the results were in the contrary direction—nine *versus* seventeen out of forty-four. This is a difference low in statistical significance, but its direction suggests that the prediction was wrong: that to express ambitiousness through the accumulation of possessions, as distinct from personal strivings, demands a more expansive outlook than can be expected of the anxious.

(14) The remaining items come from the interview. The first question invited the child to talk about his common daydreams. It was here predicted that anxious children would be of more stunted imagination and, therefore, unable to answer this question. The difference found (ten *versus* eight out of forty) is not significant.

(15) A further question asked: "On the whole, do you find that when you want something very much you usually get it?" followed by "What usually makes this 'come off'?" and "What is usually the reason if it doesn't 'come off'?" A count was made of the references to personal earnings and savings here, with the prediction, as in items 3 and 4, that they would occur more frequently with the 'anxious' group. The results were in the predicted direction (twenty-one and fifteen out of forty) but not significantly so.

(16) One question asked the reasons for people 'doing well,' offering a choice of hard work, luck and influence. It was predicted that more 'anxious' children would choose 'hard work,' but, in fact, this was a common stereotype and the incidence was high in both groups—thirty-one and thirty out of forty.

(17) The last item refers to the question "If you could have any job at all in the world and didn't have to worry about money or anything else, what would you choose?" As in item 14, it was predicted that more of the 'anxious' group would be unable to answer. The difference was not significant, the frequencies being seven and five out of forty.

Finally, the total of signs shown by each child was calculated, using quartiles for items 1 and 2 and omitting items 10, 11, 13, and 16 for reasons that will be obvious from the discussion of these items. In only seven pairs did the control score higher than the 'anxious' child, and the application of the *t* test for paired observations to the total data yields a *t* closely approaching the 1 per cent. level of significance.

#### V.—CONCLUSIONS.

That the examination of single signs of anxiety has yielded little of significance is not very surprising in view of the following two points:

(a) The 'anxious' group here was not composed solely of extreme scorers; thus, a positive goal discrepancy score almost down to the median was accepted to make a group large enough for further study.

(b) The material analysed was not originally intended for the purpose of uncovering anxiety; incidental signs, only, are the object of the search.

Nevertheless, taken together, the differences become significant and suggest the following summary of the 'anxious' child; that he or she is somewhat constrained in approaching life, concerned to find and retain security and to avoid situations which might arouse anxiety. This, perhaps, points the way to the design of further, more pertinent enquiries. Finally, it is not possible, in the present study, to say anything further about school streaming and its relationship to personality.



## VI.—REFERENCES.

- <sup>1</sup> BALKEN, E. R., and MASSERMAN, J. H. (1947). The language of phantasy: III—The language of the phantasies of patients with conversion hysteria, anxiety state and obsessive-compulsive neuroses. In TOMKINS, S. S. (Ed.) *Comparative Psychopathology*. Cambridge Mass.: Harvard U.P.
- <sup>2</sup> BROWN, J. (1956). A modified significance test for the difference between two observed proportions. *Occup. Psychol.*, 30, 169-174.
- <sup>3</sup> EYSENCK, H. J. (1947). *Dimensions of Personality*. London: Kegan Paul, Trench and Trubner.
- <sup>4</sup> EYSENCK, H. J., and HIMMELWEIT, H. T. (1946). An experimental study of the reactions of neurotics to experiences of success and failure. *J. Gen. Psychol.*, 35, 59-75.
- <sup>5</sup> FRANK, J. D. (1935). Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 47, 119-128.
- <sup>6</sup> MILLER, D. R. (1951). Responses of psychiatric patients to threats of failure. *J. Abnorm. Soc. Psychol.*, 46, 378-387.
- <sup>7</sup> SEARS, P. S. (1940). Levels of aspiration in academically successful and unsuccessful children. *J. Abnorm. Soc. Psychol.*, 35, 498-536.
- <sup>8</sup> SEARS, P. S. (1941). Level of aspiration in relation to some variables of personality: clinical studies. *J. Soc. Psychol.*, 14, 311-336.
- <sup>9</sup> VERNON, P. E. (Ed.) (1957). *Secondary School Selection*. London: Methuen.
- <sup>10</sup> WEST, V. V. H. (1958). "The Best Moment of My Life" as seen by 900 School Leavers in England. *Revue AIOP*, 2, 16-18.

*Manuscript received, March 11th, 1959.)*

# A NOTE ON THE RESPONSES MADE BY SECONDARY SCHOOL CHILDREN IN THEIR DEFINITIONS OF WORDS

By DONALD G. BURNS

(*Education Department, University of Leeds*)

**SUMMARY.** Investigations into the nature of children's definitions commonly imply some classification based on the formal characteristics of the definition (definition by synonym, by class, by description and so on). Examination of a large number of definitions from school children in Leeds suggested that it was impossible to group definitions with confidence in categories of this kind and that any attempt to assess their value should be based on criteria of a psychological kind.

Definition by emotional tone appears to be one of the most rudimentary types of definition and though such definitions have frequently very little social value, they evidently occupy an important place in a psychological scale of values.

Definition that associates meaning with a specific situation often appears superior to definition by emotional tone because it suggests a more socialised and mature concept: but within this category, many levels may be noted, ranging from the proto-class to the purely emotional.

Definitions that are generic in type, seem to reveal greater social maturity and would appear to be superior logically and psychologically to definitions in terms of specific contexts or emotional tone. Four kinds of definition of this type are distinguished.

Various levels of definition are to be found within each of these main categories and it is suggested that the level of definition and its category should be considered separate criteria in any scale of semantic values. In general, the category indicates the extent to which the definition may be logically adequate and the level, the extent to which it reflects maturity of understanding. A good definition is satisfactory by either criterion but any definition that is less complete than this, should presumably, be judged by a criterion of conceptual as well as logical adequacy and in many situations (e.g., psychological tests), the former may be more useful than the latter.

## I.—INTRODUCTION.

THE general characteristics of definition are that it should state the class to which an object belongs and the features that distinguish it from other objects in the class. Children's definitions often fail to satisfy these criteria, not merely because their knowledge is limited, but apparently, because their concepts are immature; and if this is so, there seems to be a strong case for assessing definition-answers by criteria of a psychological as well as a logical kind. The following note records some of the psychological considerations which were taken into account in marking children's definitions from this point of view, and also some of the difficulties that seem to be implicit in any gnomic classification of definitions, during an enquiry that has been completed recently in Leeds.

A large group of senior pupils in the secondary modern schools of the city was asked to give written definitions of a number of words as part of a study of vocabulary achievement in the English language. Some 4,150 words were selected from *The Teacher's Word Book* and other sources and were then graded using children's definitions as a criterion. The arrangements for the enquiry were made through the courtesy of the Chief Education Officer in Leeds, and some 3,600 pupils participated from twenty large secondary modern schools, all of whom were in their third year of study. These pupils appeared (from empirical enquiries made in the schools about their socio-economic status) to be representative of all children attending the secondary modern schools of the L.E.A.

The I.Q.s. of children in the group ranged from 70 to 115 (according to a Moray House test administered at 10 plus) and as the enquiry covered virtually all children in secondary modern schools at that time (1955), the results may be assumed to be fairly representative of all children at that stage of education in the secondary modern schools of the county borough. Further details of the method of enquiry and the basis on which words were selected have been given in *Occasional Paper No. 1* (to be published by the National Foundation for Educational Research). Each pupil was asked to give written definitions of a proportion of words in the list which were presented in a brief context—the simplest way of showing them in use—and the definitions were then given a mark on a three-point scale (1,  $\frac{1}{2}$  or 0).

## II.—FORMS OF DEFINITIONS USED BY CHILDREN.

Many investigations refer to characteristic differences in the quality of response between older and younger children when they are defining simple words. According to Terman and Merrill, "an adult gives a response which classifies a common object, but also isolates it by excluding other objects which do not belong to the class; both a six-year-old and a sixteen-year-old know the meaning of *orange*—'It is round and has seeds' or 'it is a citrus fruit,' but their thinking is characteristically different" (Terman and Merrill, p. 303). Piaget claims that the thinking of the young child is based on intuition and personal schemas of analogy and that his interpretation of experience, verbal or otherwise, is characteristically egocentric; "childish egocentrism seems to us considerable only up to about 7 or 8, the age at which the habits of social thought are beginning to be formed" (Piaget, p. 127).

Burt, too, found that definitions by description or in terms of the class to which a thing belongs are characteristic of older children (Burt, 1947, p. 42); and Feifel and Lorge confirmed this conclusion in an analysis of the responses of some 900 children (aged 6-14) to the Stanford-Binet vocabulary test. These investigators found that younger children responded more frequently with explanations in terms of use and description and tended to interpret words in a concrete sense whereas older children responded more frequently with synonyms and tended to stress the abstract features of word-meanings; the commonest of all forms of definition among children aged 9 or over in this enquiry was definition by synonym. Baranyai, in a recent article in this *Journal*, quotes further evidence on the verbal comprehension of children (aged 8-10) as shown in oral definition, which is interpreted to show a progressive development in children's thinking from the concrete to the abstract.

The evidence from all these sources suggests that a close parallel may be drawn between children's physical age and their ways of thinking or interpreting words, and one would accordingly expect the definitions of a large and representative group of children such as that examined in Leeds to provide further evidence of these characteristic ways of interpretation. A scrutiny of the large number of definitions received in this enquiry showed, however, that children at 13-14 years of age use many different forms of definition to explain the meaning of a word, depending apparently on their knowledge of the word and their understanding of what was required of them. Words they did not know well were commonly defined in simpler and less 'mature' ways (i.e., in terms of use or relation to other things) than words that were obviously more familiar, and it was difficult to discern evidence of a preference for any one form of definition over others.

The following examples, grouped according to the logical categories suggested by Robinson (Definition, 1951), have been quoted to show the variety of forms



that children's definitions may take at this age, and incidentally, the difficulties that may arise in distinguishing between them except in the most empirical way.

*Definition by synonym.*—Many children sought to define a word by using a synonym, according to Robinson a form of definition that is most successful in defining names of organic kinds. Definitions of this kind have obvious limitations which would seem to bear out the common doctrine that 'there are no synonyms':

They cross the <i>abyss</i>	canyon
The car has no <i>accessories</i>	spare parts
A pair of <i>antennae</i>	horns of insects
He serves in the <i>artillery</i>	gun army
He received a <i>bonus</i>	an extra wage during the year.
Covered with <i>bristles</i>	hard hairs
He fell on the <i>bunk</i>	bed on ships
News of the <i>catastrophe</i>	terrible happenings
A group of <i>cavaliers</i>	Royalist supporters
A <i>carbon</i> deposit	soot
The dance was <i>cancelled</i>	called off
Made of <i>chalk</i>	kind of white rock

*Definition by analysis.*—Others sought to define a word by analysing it :

Her <i>accession</i> to the throne	coming to the throne
An outstanding <i>accomplishment</i>	something to be proud of
Saved from <i>bankruptcy</i>	when you've to pay a lot of money and you haven't got it.
He <i>advertised</i> the sale	put a notice telling people to come
An <i>alluring</i> picture	draws your eyes
He opened the <i>bureau</i>	building from which you find out things
Water in the <i>carburettor</i>	it holds the petrol as it comes from the tank
<i>Approved</i> by the censor	anyone who sees a film before it goes round
<i>Besieged</i> for months	held in a place without supplies.

*Definition by synthesis.*—Some children attempted to define a word by stating its relation with other objects that are commonly associated with it, a method which seemed successful provided the synthesis was close and unambiguous :

They gathered <i>acorns</i>	from an oak-tree
He fell on the <i>bunk</i>	it was two-tiered
Suffering from <i>concussion</i>	where you lose your memory

*Definition by implication.*—Others implied that the concept was already familiar to the reader and gave examples of things to which it might refer. The logical inadequacy of definitions of this type is evident and inevitably prompts the question "Under what conditions can a partial statement be assumed to reveal an adequate knowledge of the concept?"

Interested in <i>agriculture</i>	crops and animals
Infected with <i>bacteria</i>	microbes and germs
Here is your <i>baggage</i>	trunks and cases
Infested with <i>bugs</i>	insects and flies
Blamed for <i>carelessness</i>	for spilling something
<i>Catering</i> for hundreds	making tea, etc.
A crop of <i>cereals</i>	crop of wheat, barley and corn
A <i>superstitious</i> belief	she would not walk under a ladder

*The denotative method.*—In some answers, the definition showed the word in use in a different context. This is clearly an alternative form of definition by implication :

Where is your <i>accomplice</i> ?	the thief and his accomplice got away
A sharp <i>angle</i>	thirty degrees is a sharp angle
An <i>annual</i> event	it was the annual fete-day on Saturday
She accepted the <i>bouquet</i>	bouquet of flowers
He has been <i>bitten</i>	bitten by a snake or a dog

*The ostensive method.*—One other method which was observed in some answers seems to depend more or less directly on the learner's experience of the example. This also appears to be another form of definition by implication :

They <i>reacted</i> favourably.	if someone stuck a pin in you, you would react
Within my <i>recollection</i>	when you remember
A number of <i>restrictions</i>	things you must not do
Accused of <i>selfishness</i>	all for yourself
A difficult <i>target</i>	something you aim at

The difficulty of distinguishing between these different forms of definition is evident. Some of those classed above as 'definitions by synonym,' for example (because they are attempts to define by description or in terms of the class to which a thing belongs), might equally be considered examples of definition by implication. Is *soot* a synonym for *carbon* or an example of it? And are *antennae* exemplified or paraphrased by the definition 'horns of insects'? Similarly, the form of definition offered for *cereal* (a crop of wheat, barley and corn) and here classed as definition by implication, might be judged an example of definition by synonym. In such cases—and there were many of them—it was impossible to group definitions with confidence in any of the logical categories mentioned above, not did it seem that any particular significance could be assigned to the phrasing of a definition, whether this was in the form of a synonym, description or statement of the class to which a thing belonged. The only alternative in the present enquiry was to attempt to frame more detailed criteria of a psychological kind and the possibility of formulating these is considered in the next section.

### III.—STAGES IN THE DEVELOPMENT OF A MATURED DEFINITION.

Most psychologists would agree that the young child tends to apprehend the world as dynamically related to and dependent on himself and Allport writes that "between the primitive and the child and his environment, there exists a raw or syncretic relation illustrated by fusions between dreams or imagery and perceptions of outer reality . . . (and) it is only in the most highly developed individuals that one finds a sharpening of the polarity between subject and world" (Werner, 1958). Werner, in developing this thesis, has proposed that certain types of mental organisation or structure are characteristic of particular levels of development and may be deduced from parallels observed in the form and structure of the thinking, conceptualising and perceiving of primitives, children and psychotics. This too, is essentially the view of Piaget, though many would disagree with his conclusion that egocentric and syncretic modes of thinking are characteristic of a particular stage of childhood. Indeed, it seems probable, as Vernon suggests (Vernon, 1955), that such modes

of thinking represent stages in the development of each class of ideas which may occur at any age. But the general conclusion seems clear: the child's concepts develop gradually and are continuously being modified as a result of experience so that what were originally 'raw syncretic relations' become progressively more informed and more sharply determined. Defining words is a function of this process of socialisation just as much as other forms of mental behaviour (e.g., making logical deductions), and an assessment of the value of a definition is, in effect, an assessment of the extent to which the child has assimilated the social implications of the word.

There are marked differences, however, in the extent to which a word can be defined with accuracy. Some have a meaning which can be communicated, and therefore assessed, with some precision because the senses in which they are used are closely prescribed by custom; examples of these are the large number of substantives and verbs which can be defined ostensively like *boot*, *door*, *walk*, and *ride*. Others have developed meanings which are frequently less precise because the reference is largely determined by social contacts, and definitions of these cannot be assessed with the same confidence; examples are *amenities*, *dingy* and *elegant*. And there are others for which the field of reference may be almost wholly determined by personal experience (*beautiful*, the *establishment*, *image*) and interpretation then becomes a matter of individual judgment. It follows that children may develop a satisfactory knowledge of some words much more quickly than others and that their familiarity with more complex concepts may have to be assessed by some predetermined standard of definition. An arbitrary standard of this kind seems necessary, for example, in assessing definitions of abstract words and words used metaphorically.

One of the least sophisticated modes of definition and possibly one of the most rudimentary, is definition by emotional tone. Definition of this kind is readily understandable in early life since, at that stage, all the child's articulate sounds are the expressions of feelings or conational attitudes. Meumann and Stern have suggested that the earliest substantives of child language are very far from denoting concepts but rather express commands or desires: and Valentine has shown that young children frequently associate new words through some emotional tone which is felt to be common to them before they learn to discriminate attributes more exactly. One of Valentine's children, for example, appeared to use the word *hot* as meaning either general unpleasantness or both very hot and very cold (Valentine, page 414). O'Shea quotes the case of an older child (aged 4) who began to use the word *imagination* in the sense of any undesirable quality in objects, apparently because of the way in which his elders used it in his presence (cf. the phrase "he has a very lively imagination") (O'Shea, p. 138). Werner, who finds evidence of rudimentary and even primitive ways of thinking in the interpretation of meaning in emotional terms ("physiognomic perception") cites many examples of similar usage among older people: the tripod of a camera may be described as 'proud' when it stands stiff and 'sad' when it leans at a precarious angle; just as a figure with sharp corners may be described as 'cruel' (Werner, p. 256). Indeed, Osgood and his collaborators have suggested that the emotional tones that a word arouses in different individuals may be sufficiently significant to be used as one of the three basic dimensions of meaning, the other dimensions being potency and activity (Osgood, Suci and Tannenbaum). Definition by emotional tone is evidently not a form of definition used only by young children (though it is typically egocentric in Piaget's sense of the term) and many examples of definition of this kind were found among the answers of children aged 13 and 14 in Leeds.



An <i>aristocrat</i> by birth	comedian
He <i>abused</i> her kindness	scorned
An <i>advantageous</i> position	good
An <i>austere</i> man	wicked
He is a <i>bolshevik</i>	untidy person
He became <i>Lord-Chancellor</i>	someone clever
A <i>chronic</i> disease	very bad
A well-known <i>communist</i>	powerful

Definition at this level seems to be a pre-cognitive response, and identification apparently occurs between any elements that share the same emotional context. This might explain many of the instances of definition by antonym which occurred in children's replies in Leeds. A confusion between, e.g., *dusky* and *colourless*, may be due to association on an emotional rather than a cognitive level and is certainly not unlike the confusion of Valentine's child between *hot* and *cold*:

A <i>dusky</i> skin	colourless
This is a <i>harmful</i> drug	won't hurt
In a <i>hilarious</i> mood	in a bad mood
The place is <i>inhabited</i>	abandoned: a deserted place
An <i>immigrant</i> family	to go abroad
Of <i>inferior</i> quality	best: better quality
I am <i>indebted</i> to you	I don't owe you anything
An <i>irresponsible</i> person	take charge of things
He <i>loathes</i> this food	likes
She was my <i>predecessor</i>	successor
Living in <i>poverty</i>	luxury
He insisted on <i>punctuality</i>	lateness
She <i>preceded</i> me	came after

A clear and easily recognised step in the development of a more socialised concept appears to be revealed when meaning is identified with a specific context:

I <i>conjure</i> you to give up	magic: magician
A closed <i>vehicle</i>	black maria

Definition which seeks to locate meaning, as here, in terms of some finite and cognitive experience, reflects a notion of the concept which is potentially superior to definition by emotional tone, however incomplete it may be.

This also appears to be a form of definition to which children regress when they are unable to apprehend or express wider concepts. Watts observed from a survey of the errors of children aged 11 and upwards in a proverbs and idioms test, that when they were in doubt or ignorant of the meaning of a proverb, they selected a sentence which corresponded as closely as possible to the literal meaning of the proverb and this is analogous to definition in terms of a specific context, which is also an attempt to identify generic meaning with the literal one. He notes also that "In seeking to select what they regarded as the literal meaning of a proverb, children seemed to look first for a sentence containing some striking word which they remembered from the corresponding proverb" (Watts, p. 215). Thus, "One man's meat is another man's poison" was frequently identified by children as meaning the same as "Beware of eating poisoned meat when dining with a stranger."

Watt's procedure is very similar to that reported by Piaget though his interpretation of the results is quite different. In many cases, the children

questioned on the meaning of proverbs by Piaget at the age of 9, 10 or 11, did not understand them "in the least" though they evidently thought they did because they asked for no supplementary explanation of their literal or hidden meaning (Piaget, p. 129): Matt connected the proverb "so often goes the jug to the water that in the end it breaks" with the sentence "As we grow older, we grow better" and when asked why these two sentences meant the same thing, said "Because the bigger you grow, the better you are and you grow old" (i.e., the jug is compared to a child simply because both grow older). Piaget considers this to be evidence of the syncretistic reasoning which is characteristic of all children at this age; and Watts interprets it as an association between two fundaments which have some superficial or literal identity.

It seems probable that the degree of syncretism revealed by children's responses in situations of this kind depends very much on the complexity of the material. The more concrete examples offered by Watts, for example, may have facilitated the identification of the proverb with a literal interpretation and this is a procedure which seems to be analogous to definition in terms of a specific instance.

A number of the definitions given by children in Leeds seem to exemplify this regression from a broader concept (which they were unable to explain) to a narrower definition based on some specific experience. The definition was apparently framed to express a generic concept of the word but was qualified or explained by the addition of details which make it analogous to other forms of definition with a specific context:

Along the <i>embankment</i>	place where trains run
A feeling of <i>guilt</i>	feeling of dishonesty
They paid <i>homage</i>	tributes for the last time
He received a <i>pension</i>	reward for something you have done
He became <i>premier</i>	first prime minister
Classified as a <i>quadruped</i>	has three legs
To <i>safeguard</i> our health	somebody who guards people
It contains a <i>supplement</i>	something free—a knitting pattern
Serving as a <i>trooper</i>	man who bails from plane

All definitions that are generic in kind suggest a more fully developed understanding of the concept than definitions in terms of a specific context and various forms of generic definition were noted in children's answers.

One of these was the plain definition of the word by stating its genus (*Orange* is a citrus fruit). The following are examples of definitions of this type, all of which are evidently incomplete because they fail to state the distinguishing marks by which a concept differs from other members of the same class:

This will not <i>kindle</i>	burn
The curtains are <i>ornamental</i>	for show
A <i>murderous</i> blow	he killed him with a blow
He found the <i>parchment</i>	paper with some important thing on
Living like a <i>peasant</i>	poor man
The <i>origin</i> of the quarrel	cause
He <i>speculated</i> on the matter	look into
He was <i>summoned</i> yesterday	found guilty
<i>Transparent</i> material	invisible
<i>Wielding</i> a sword	swinging

In other definitions, the class concept may seem to be implied but is not made explicit. This may be because children were unable to express their knowledge in words or simply because they felt they were only expected to

show they could identify it in an approximate way. The fact that children's awareness of meaning may outrun their ability to express it has been noted elsewhere. O'Shea writes, "I asked H., at seven to define loveliness. I can tell from her response when the term is used, in conversation or reading, that she understands it in its fundamental reference at any rate, and she can use it quite effectively; but she cannot state formally just what meaning it has for her. "I know but I cannot tell," she says . . . if I press H. for some statement about *loveliness*, she will fall back upon the simpler term *lovely*—"when a thing is lovely, it has loveliness." (O'Shea, p. 146). Older children must face the same problem and may, as Werner and Kaplan have suggested, attempt to bridge the gap between the general and the specific by using phrases such as 'a kind of,' 'like a,' or 'sort of.' Such formulae seem to argue an awareness of the need for a generic definition, and to this extent, may be held to reveal a more mature concept than any form of specific definition but they can hardly be considered the equivalent of a class concept. :

A <i>compartment</i> to himself	kind of small room
They ran the <i>blockade</i>	kind of barricade
A closed <i>vehicle</i>	a car or something
She entered the <i>convent</i>	kind of school
Shaped like a <i>cone</i>	like an ice-cream

Other definitions that were phrased in this way, included some mention of the characteristic features of the concept :

The <i>boar</i> charged	like a pig with tusks
She wore a <i>coronet</i>	jewelled hat like a crown

A fourth form of generic definition—and one that was generally most satisfactory from a logical point of view—was that which stated the class as well as the distinguishing features of the concept :

An <i>almond</i> cake	cake made with nuts called almonds
An <i>inquisitive</i> person	for ever questioning
She often <i>flirted</i>	made love for fun
He <i>scanned</i> the horizon	examined carefully
He <i>cantered</i> into the yard	gallop gently
She <i>sipped</i> her tea	drank a little at a time

There were many words, however, that children of 13 and 14 could hardly be expected to define adequately *per genus et differentiam* because they lacked the maturity or specialised experience of the adult and here, some less exacting form of definition must probably be considered adequate. Burt has noted that few children are able to give more than an approximate definition of words like *kindness* and *justice* before the age of 12 or 13 (Burt, 1955, p. 352), and younger children tend to define such complex concepts, as might be expected, in terms of specific instances. "Punctuality means coming to school at a stated hour, discipline means being ready to obey definite rules on given occasions, truth means saying precisely what you have seen and heard and done and so on" (Watts, p.26). Definitions at this level are, perhaps, evidence of a habitual use of the words in these situations rather than of any inability to deal with abstract concepts as such; and the apparent inability of older children to give adequate definitions of more complex concepts can be similarly explained. Often their acquaintance with a word is limited to two or three contexts, which is a very small basis for determining genus and species. In such cases, any definition



that states or implies the generic relationship and exemplifies it in a particular way, would seem to be satisfactory, despite its logical limitations :

The <i>accusation</i> is unjust	you accuse somebody when you say— you've done it
The city <i>corporation</i>	people to whom we pay rates and in return, they clean the streets
He showed great <i>friendliness</i>	kindness to help

#### IV.—THE INTERPRETATION OF WRITTEN DEFINITIONS.

In a review of written definitions, there must be many aspects of conceptual development that escape analysis owing to the form in which the definitions were given. It is difficult for example, to discriminate with confidence between varying levels of the wide range of emotive definitions when these are given in written form. Other aspects of conceptual development such as the 'plural concept'—the genetic predecessor of the universal concept, according to Stern—have been omitted from the discussion because they seemed to be of marginal interest and could only be adequately considered in a much longer paper. The limitations of any study of conceptual development based on written definitions are indeed self-evident and the chief interest of the present survey may lie in the evidence it provides of concepts at virtually every level of development among a group of older children.

According to this evidence, it would seem that the most rudimentary type of definition assumes the form of an equation between meaning and emotional tone (e.g., *aggressive* is bad-tempered, *austere* is wicked). All definitions that have an emotive rather than a cognitive basis have been grouped in this general category but the value of the specific definition will evidently depend on its social relevance. Some (e.g., definition by antonym) may be judged to be more relevant than others (e.g., the syncretistic interpretations quoted by Piaget) and a few may have passed into current use (e.g., The *coarseness* of the cloth . . . thickness : A *chaste* life . . . hard, stern). Whatever the social use of such definitions, they clearly have an important place in a psychological scale of values.

Definition that associates meaning with a specific context appears to be markedly superior to definition by emotional tone because of the evidence it should offer of a more socialised and mature concept ; but here again, many levels of definition were noted ranging from the proto-class concept to the purely emotive. Some suggest a broader concept and imply a regression to a narrower interpretation by quoting some specific experience (e.g., *Pension*—reward for something you have done), some are practically valueless because the specific context is incongruous (e.g., the young boy's use of *imagination*) and others that are based on a superficial resemblance between fundamentals or some form of syncretistic reasoning, are psychologically indistinguishable from the emotive definition.

Definitions that are generic in type, seem to reveal greater social maturity and are logically and psychologically superior to definitions in terms of a single context or an emotional tone. Several kinds of definition of this type were observed :

- (a) Definition stating the class concept.
- (b) Definition implying the class concept.
- (c) Definition stating class and significant characteristics.
- (d) Definition implying class and stating characteristics.

A number of definitions that stated the class concept seemed to be satisfactory, no doubt because the synthesis between word and class-concept was close (i.e., they were virtually synonymous); *befell*—happened, *needless*—unnecessary, *slaughter*—killing. The difference between these and definitions that implied the class concept by using phrases of the kind 'sort of,' 'like a,' 'a kind of,' seems to be only one of degree.

For many concepts however, definition by class and characteristic is the only satisfactory form of definition (*Cantered . . . galloped gently, sipped . . . drank a little at a time, scanned . . . examined carefully*). Logically adequate definitions of this type for complex concepts can hardly be expected from children of 13 or 14; and it seems reasonable to consider any definition satisfactory that relates the class concept to a specific instance, assuming of course, that both concept and instance are appropriate (*Corporation . . . people to whom we pay rates and in return they clean the streets*).

Varying levels of definition may accordingly be found within each of the main categories postulated by this analysis and it would seem useful to consider the level and the category (as here defined) as separate criteria in any scale of semantic values. In general, the category indicates the extent to which the definition may be logically adequate and the level, the extent to which it reflects maturity of understanding; and a whole range of definition may be offered for any word by either of these criteria. A good definition is satisfactory by either criterion because it defines *per genus et differentiam* and by the same token, reveals a mature concept; but any definition that is less complete than this must presumably be judged by a criterion of conceptual maturity as well as one of logical sufficiency.

#### V.—BIBLIOGRAPHY.

- BARANYAI, E. H. (1958). Verbal comprehension in Hungarian children of 8-10 years. *Brit. J. Educ. Psychol.*, 28, 262-265.
- BURT, C. (1947). *Mental and Scholastic Tests*, 2nd ed. Staples.
- BURT, C. (1955). *The Subnormal Mind*, 3rd ed. Oxford Univ. Press.
- FEIFEL, H., and LORGE, I. (1950). Qualitative differences in the vocabulary of children. *J. Educ. Psychol.*, 41, 1-18.
- MEREDITH, G. P. *Dimensions of Ambiguity*. (Publication pending.)
- O'SHEA, M. N. (1907). *Linguistic Development and Education*. New York: Macmillan.
- OSGOOD, G. E., SUCI, G. J., and TANNENBAUM, P. H. (1958). *The Measurement of Meaning*. Univ. Illinois Press.
- PIAGET, J. (1932). *The Language and Thought of the Child*. Routledge and Kegan Paul.
- ROBINSON, R. (1950). *Definition*. Oxford Univ. Press.
- TERMAN, L. M., and MERRILL, M. A. (1937). *Measuring Intelligence*. Harrap.
- THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word-book of 30,000 Words*. New York: Teachers' College, Columbia.
- VALENTINE, C. W. (1950). *The Psychology of Early Childhood*. Methuen.
- WATTS, A. F. (1944). *The Language and Mental Development of Children*. Harrap.
- VERNON, P. E. (1955). The psychology of intelligence and G. *Quart. Bull. Brit. Psychol.*, No. 26.
- WERNER, H. (1957). *Comparative Psychology of Mental Development*. New York: International Univ. Press Inc.
- WERNER, H., and KAPLAN, E. (1952). *The Acquisition of Word-meanings—A Developmental Study*. Child Development Pubs., North-Western University.

# THE PERFORMANCE IN ENGLISH LANGUAGE, AT 'O' LEVEL, OF A SAMPLE OF UNIVERSITY STUDENTS

By E. A. HEWITT

(*Durham Colleges Education Department, University of Durham.*)

**SUMMARY.** Results in the 1953 and 1954 Cambridge examinations in English Language, at 'O' level, have been obtained for 1,423 boys and 882 girls, from 197 schools, who were subsequently admitted to universities. Significant differences are shown in the grades awarded to boys and girls, and to arts and science students, of comparable age. Linguists achieved significantly better marks than non-linguists, but there is no significant difference between the non-linguists among the arts men and the men who specialised in the 'pure' sciences.

The failure rates for boys taking the paper for the first time were 1 in 9 for scientists, 1 in 15 for arts students. For girls the corresponding rates were 1 in 42 and 1 in 27. The male scientists who had won open awards to universities were not inferior in English, at 'O' level, to arts students in general, though they were inferior to the arts men who won open awards.

Among the arts men, those who went to Oxford, Cambridge or London gained significantly better marks than those who entered other universities; a similar difference is found in respect of the men scientists, and of the women arts students.

## I.—INTRODUCTION.

It is generally accepted that the written English of university science students is inferior to that of arts students, and the inferiority appears usually to be attributed to specialisation at university and in the last years at school. The chief remedy so far employed has been the provision of more time for scientists to pursue English studies in the VIth form, though some university dons are now giving their scientists linguistic training.\* It has seemed to the present writer that the written work of some arts students and of more scientists contains faults that one would hope not to find in the work of the average grammar school pupil at 'O' level, and that, before specialisation begins, the prospective arts student writes better English than the pupil who later specialises in one of the sciences. In general, however, headmasters and teachers with whom the writer has discussed the matter have expressed the view that at 'O' level the prospective scientist is as good at English as the future arts student.

## II.—SCOPE OF THE ENQUIRY.

Results in the Cambridge Syndicate's 'O' level examination in English Language have been obtained for 1,423 boys and for 882 girls who took the examination for the first time in 1953 or 1954, and who have since been admitted to universities—or, in the case of some of the men who did their national service first, have been offered, and have accepted, places for 1959. The writer is indebted to the Council of the Durham Colleges, who financed the distribution of a questionnaire to more than 400 schools; to the headmasters and headmistresses who replied to it; and to the Cambridge Syndicate, whose officials were most helpful at every stage of the enquiry. The questionnaire asked for candidates' names, examination result and date, together with candidates' ages, the university to which they were admitted, and whether they followed a degree course in arts or in science. It was suggested that such further details as scholarship awards, subjects in which candidates were specialising at university, and whether they had taken the 'O' level Language paper more than once, would be of interest. Naturally, in view of the time required for looking up

\* See, for example, *Times Ed. Supp.*, Oct. 17th, 1958: Towards the broader view.



records, schools with large numbers of university entrants tended to be more selective in their replies to the questions. In some cases pupils' ages were not given; in others, specialist subjects, or scholarships, were not mentioned; the universities to which eight boys had gone were not named. These omissions restrict, sometimes severely, the size of the sub-samples in the tables given in the Appendix.

The official Cambridge list for 1958, showing 526 centres in England where candidates sat for the Syndicate's papers, was used in distributing the questionnaire. Some schools and institutions appearing in this list entered pupils for certain Cambridge papers, but not for the English Language paper. Others were new schools, or, for various reasons, had only begun to enter pupils for the Language paper after the relevant years 1953 and 1954, and the list did not, of course, include any centres which may have transferred their allegiance from Cambridge after the latter year. No approach was made to 112 centres where, because of the type of school or institution, or because of the small numbers on the roll, it seemed clear that very few pupils could have proceeded to universities. Table 1 gives the numbers and types of schools and other educational establishments in the 1958 list, of those which answered the questionnaire, of those which did not, and of those which were not approached. Replies were received from 73 per cent. of the schools and institutions to which the enquiry was addressed.

TABLE 1  
NUMBERS AND TYPES OF EDUCATIONAL INSTITUTIONS ENTERING CANDIDATES FOR THE  
CAMBRIDGE 'O' LEVEL PAPER IN ENGLISH LANGUAGE.

		Maintained Grammar and Direct Grant * Schools	Independ- ent Schools	Sec. Mod. and Sec. Tech. Schools and other Institu- tions	Totals
Institutions providing details of ex-pupils at universities	Boys	51	13	3	67
	Girls	54	40	1	95
	Mixed	27	7	1	35
Institutions rendering nil returns .....	Boys	6	10	25	41
	Girls	14	26	7	47
	Mixed	6	1	10	17
Institutions unable to fill in questionnaire .....	Boys	11	1	—	12
	Girls	4	—	—	4
	Mixed	—	—	—	—
Institutions not replying to enquiry .....	Boys	17	14	5	36
	Girls	6	41	—	47
	Mixed	12	1	—	13
Institutions to which the enquiry was not addressed	Boys	—	10	43	53
	Girls	—	30	14	44
	Mixed	—	4	11	15
Totals .....	Boys	85	48	76	209
	Girls	78	137	22	237
	Mixed	45	13	22	80
All Types		208	198	120	526

\* Includes 'voluntary controlled' and 'voluntary aided.'

The boys in the sample would be roughly equivalent numerically to a 5 per cent. sample of English boys entering universities in the years 1955 and 1956, and would account for perhaps a similar percentage of the boys who won State Scholarships. The girls constitute a 6 per cent.—7 per cent. sample, and include rather more than 7 per cent. of the girls who won State Scholarships. The chief defect in both samples is probably that London and the North of England are very thinly represented.

### III.—THE DATA.

During the period 1953-1954, the Cambridge paper in English Language retained the same form. An essay on one of nine topics, and a precis, were compulsory; the remaining questions, from which two had to be selected, dealt with such matters as word and phrase equivalents in given contexts, punctuation, direct and indirect speech, and the re-writing of a faulty prose passage. There were no questions dealing with formal grammar or clause analysis. The original marks awarded are no longer available, for the results were published in terms of nine grades, grade five being the lowest passing grade. The mark intervals corresponding to the grades are known, but the grades do not cover equal ranges of marks, and the highest and lowest grades are, in effect, open-ended, so that the means for various groupings of pupils cannot be calculated, though medians can be. While the grades are designed to be comparable from year to year, the range of marks within any one grade is not necessarily the same from one year to the next, so that medians should be calculated for each year separately. In practice, these difficulties have been of little consequence, for the distributions of candidates among the different grades are of more interest than means or medians. The distributions for various groups of pupils are shown in the Appendix, and differences between groups have been calculated in terms of  $\chi^2$ .

### IV.—RESULTS.

(a) *Comparison between arts and science students.*—The terms 'arts' and 'science' are used here to refer to groups of subjects, not to university faculties. Law, and Economics, for instance, are classified under 'arts'; 'science' includes Mathematics, Engineering, Medicine, Dentistry, Agriculture and so on. In a few cases, classification presented difficulty, but the border-line decisions have not materially affected results.

In each year, and for both boys and girls, the median of the prospective arts specialists is roughly three marks higher (on a scale running in theory from 0 to 100, though the range of marks between grades 1 and 9 is only 35) than that for the scientists. The English grades for boys are shown in the first contingency table in the Appendix. There are proportionately fewer scientists in the upper grades, more in the lower. The difference between the distributions for arts and science students is highly significant ( $\chi^2$  with 5 degrees of freedom = 65,  $P = .001$ ). Among the girls, too, there is a marked difference between the gradings of arts and science students ( $\chi^2$  with 5 d.f. = 38,  $P = .001$ ). Calculations for each age group separately provide no evidence to suggest that the differences are due to scientists taking the 'O' level paper earlier, on average, than arts students.

The boy scientists who won State Scholarships or other open awards achieved better English results than other scientists ( $\chi^2$  with 1 d.f. = 28,  $P = .001$ ); they did less well than the scholarship winners among the arts men ( $\chi^2$  with 1 d.f. = 9.7,  $P = .01$ ), but there is no significant difference between their results and those of arts men in general.

(b) *Comparison between linguists and non-linguists.*—Men reading for degrees in languages were significantly better in English at 'O' level than men studying other arts subjects ( $\chi^2$  with 2 d.f.=28.6,  $P=.001$ ). There is a tendency for those pursuing 'pure' science to get better English results than other men grouped under the general heading 'science' ( $\chi^2$  with 1 d.f.=3.86,  $P=.05$ ). But there is little to choose between men specialising in other arts subjects than languages, and those taking 'pure' science;  $\chi^2$  with 1 d.f. is .026, a value that would be equalled or exceeded in 80 per cent. of random samples.

(c) *The sex difference.*—The medians for girls who later specialised in arts and science respectively are roughly three marks higher than those for the boys. The girls' sample contains a higher proportion of pupils from independent and direct grant schools than the boys' sample does, but there is nothing to indicate that this is responsible for the apparent sex difference. Line 6 in the Appendix suggests that 16-year old girls from the maintained schools do as well in English as those from independent and direct grant schools ( $\chi^2$  with 4 d.f.=1.0, a value that would be obtained in 90 per cent. of random samples), and the difference in respect of the 15-year-olds does not reach the 5 per cent. level of significance.

Among both arts and science specialists, the differences between the distributions for boys and girls are significant beyond the 0.1 per cent. level, and separate calculations for 15-year-olds and for 16-year-olds suggest that the apparent superiority of the girls is not to be accounted for by any tendency for boys to take the examination at an earlier age than is the case with girls.

(d) *Distribution of students among the universities.*—One might anticipate that Oxford, Cambridge, and London would between them take a disproportionate number of the arts students who are most highly gifted verbally, and in fact, the distributions for these three universities on the one hand, and for the remaining British universities on the other, are significantly different, at the 0.1 per cent. level for boys ( $\chi^2$  with 5 d.f.=27) and at the 1 per cent. level for girls ( $\chi^2$  with 4 d.f.=14). The three senior universities also take a disproportionately large number of the verbally-able boy scientists—as judged by 'O' level English results; the difference between the distributions for Oxford, Cambridge, and London together, and for the other universities, is significant at the 0.1 per cent. level ( $\chi^2$  with 5 d.f.=46). When the male scientists at the three senior universities are compared with male arts students elsewhere, the difference between the distributions is no greater than would be found in 50 per cent. of random samples ( $\chi^2$  with 2 d.f.=0.9). The failure rate in the Language paper for all boys proceeding to Oxford, Cambridge and London together is 1 in 14; for those who went to other universities it is 1 in 9.

(e) *Candidates with poor results in English at 'O' level.*—Proportionately more boys than girls in the sample failed in their first attempt on the Language paper—as is the case not only with all Cambridge candidates, but with all 'O' level entrants in England.\* For the two-year sample of university entrants, the failure rates for boys are 1 in 15 and 1 in 9 for arts and science students, respectively, and the corresponding rates for girls are 1 in 42 and 1 in 27. Not all the first-time failures, of course, are to be regarded as poor material from the linguistic standpoint; one girl initially in grade eight eventually took a II (2) honours degree in English, and one boy scientist who failed in his first attempt was placed in grade one six months later. Nevertheless, in view of the fact that roughly 60 per cent. of all boys and 70 per cent. of all girls taking the Cambridge

\* See e.g., *Education in 1953*: H.M.S.O., Cmd. 9155; and *Education in 1954*: H.M.S.O., Cmd. 9521.



paper in any one year pass the examination, failure in students of university calibre is no light matter. 132 boys (9.3 per cent. of the sample) failed at the first attempt, and in 125 cases the final passing grade is known. One-third of the scientists and one-fifth of the arts students could not achieve better than the lowest passing grade when they were eventually successful. At least twenty boys required *three or more* attempts at the paper, and one scientist—now expected to take a good Ordinary degree—needed seven. All the girls who failed at the first attempt succeeded at the second.

#### V.—DISCUSSION AND CONCLUSIONS.

The sex difference shown in this enquiry may be due at least in part to the earlier maturing of girls, but it seems unlikely that the English of women university students can give rise to as much concern as that of the men, and no further reference to the girls' 'O' level performance will be made here.

The interpretation of the remaining findings must be governed by such considerations as the representativeness of the sample, the validity of the 'O' level Language paper, and the consistency of standard in the written English of individual students.

(a) *Sampling*.—Some of the factors which, in conjunction with sampling errors, might affect the apparent difference in ability in English as between arts and science specialists can be illustrated from the records of two schools each of which, in the course of two years, sent some thirty boys to universities. School A, in a south-coast resort, contributed roughly one university entrant for every fifteen boys on the school roll, and one in two won a State Scholarship or other open award; two out of every three were placed in grades 1 or 2 in English Language, and only one boy—a foreigner—failed at the first attempt. School B, serving a rural area and producing a higher proportion of scientists than School A, contributed roughly one university entrant for every six boys on the roll, one in four gaining State Scholarships; only one in ten was placed in grades 1 or 2 in English Language, and one in four failed at the first attempt. School A entered only four under-age candidates for the paper, and they were all placed in grade 1; more than half the candidates from School B were under-age, one of them, a 14-year-old, requiring two further attempts before he passed.

Such differences between schools render futile any attempt at general prescription, and emphasise the need for a wide sampling of schools in an investigation of the kind described in this paper. In spite of the variety of conditions, however, the tendency for arts specialists to achieve better results than scientists in the language paper is remarkably regular from school to school. Out of the twenty-five schools which each sent twenty or more boys to universities, the reverse trend is to be found in one, and in three schools there is no perceptible trend; in the remaining twenty-one, the arts students achieve the better results, and in each of three schools, in spite of the relatively small numbers of pupils concerned, the difference between arts and science specialists is significant at the 5 per cent. level.

It is not, of course, inconceivable that a particular examination in English should penalise scientists who, had they entered for the paper set by another examining body, might have done as well as the arts specialists. The only concrete evidence concerning the comparability of papers set by different examining bodies is that the failure rates in the Cambridge paper closely resemble those quoted by the Ministry for the papers of all nine examining bodies together. There seem to be but slender grounds for supposing that use of the Cambridge paper for the investigation reported here gives an exaggerated impression of the poor standard of the English of some prospective university entrants.

Attention has already been drawn to the fact that some of the samples are very small. Inferences from the figures given in lines 3 and 4 in the Appendix, for instance, cannot be made with the same confidence as inferences based on the first two lines.

(b) *The validity of the examination.*—Some evidence of validity has already been presented—those who specialised in languages at university level tended to get higher 'O' level marks. English Language papers at 'O' level may not be precise measuring instruments, but lack of consistency in markers, candidates and papers cannot create the distinctions revealed in Section IV above. At the worst, one would expect the paper to provide some evidence of candidates' ability to write clear English, using accepted conventions; to test width of vocabulary, grasp of sentence structure, and power to put thoughts in an appropriate order. If the essay does not promote any high degree of involvement, at least most candidates will have had much practice in the skills required for success. 60 per cent. of all boys entered for this examination pass, and it is improbable that the standard of English demanded for university entrance is any higher than that of the average 16-year-old grammar school boy. There was a failure rate of 8 per cent. among those boys in the sample who were reported to be 16 years of age when they first took the paper, and though some may have been merely unfortunate, it is difficult to believe that the English of the majority of these unsuccessful candidates could be creditable in any circumstances.

(c) *Consistency of individual performance in English.*—The available evidence suggests that a pupil's relative standing in English is reasonably stable over the first five years of secondary schooling. The correlation between 11+ quotients and S.C. or G.C.E. 'O' level English seems to be in the region of 0.5,\* uncorrected for the effects of selection, and so comparable with the correlation between a typical 11+ battery and a general academic criterion five years later. One may at least assume that those who are good at English early on tend to be good later, and that defects at 16 are likely to be associated with earlier shortcomings. Defects of long standing are not likely to be remedied quickly, and one would expect that attention given to them before the VIth form, if it were feasible, would be more effective than the provision of time and special courses later.

(d) *Main conclusions.*—Despite the possibilities of errors in sampling and weaknesses in the measuring instrument, which have been conceded above, the available evidence points to the following conclusions:

1.—The difference in ability in English between arts and science students in universities is not entirely due to the effects of specialisation; a difference can be perceived at least as early as 15.

2.—The contrast between the English of scientists and that of arts men is less fundamental than that between those reading for a degree in languages and those taking other subjects.

3.—On the whole the better scientists (judging by scholarship awards) are better at English than the other scientists, and are not significantly inferior to arts students in general.

\* The following have reported, in this *Journal*, correlations of 0.47, 0.48, 0.486, and 0.498, respectively:

Peel, E. A., and Armstrong, H. G. Symposium—The use of essays in selection at 11+: 26, 163-171.

11—The predictive power of the English composition. 21, 30-35.

Peel, E. A., and Rutter, D. The predictive value of the entrance examination. 21, 30-35.

Wiseman, S. Symposium: III—Reliability and validity. 26, 172-179.

Emmett, W. G., and Wilmut, F. S. Prediction of S.C. performance in specific subjects. 22, 52-62.

4.—The poor English of some university students does not in itself constitute a strong case against the standard of English teaching in the grammar schools. University selectors are admitting candidates who collectively present, at the age of 16 or thereabouts, a range of ability in English which is comparable with that of the whole population of grammar school boys of that age. The universities might improve the situation by taking 'O' level English results into consideration, at least when weighing the claims of border-line candidates. Complaints about the poor standard of English expression of some university students are inevitable when, as with the present sample, over 9 per cent. of the men failed in their first attempt on the 'O' level English Language paper, and 8 per cent. of the 16-year-olds similarly failed to reach the standard achieved by 60 per cent. of all boys entering for the examination.

5.—The verbally-able boys—as judged by results in the English Language paper—whether arts or science students, tend to go to one of the three senior universities. The provincial universities would probably be most affected if the standard of English required for university entrance were raised, or if, as the result of university expansion, the standard of English among university students in general declined.

## VI.—APPENDIX.

### UNIVERSITY STUDENTS' GRADES IN ENGLISH LANGUAGE AT 'O' LEVEL.

Categories of Pupils	Percentage of pupils awarded the grades indicated.						No. of pupils.
	1	2	3	4	5	6-9 (failed)	
(1) Boys reading arts .....	13.7	20.2	29.0	17.5	13.0	6.7	555
Boys reading science .....	5.1	12.1	29.1	22.8	19.9	10.9	868
(2) Girls reading arts .....	18.2	27.8	33.2	12.2	6.2	2.5	582
Girls reading science .....	7.7	23.3	31.3	22.7	11.3	3.7	300
(3) Boys with State Schols. { arts	82.5			17.5			102
or other open awards { science	64.5			35.5			169
(4) Boys—languages .....	52.9	31.4		15.7			102
Boys—other arts subjects .....	21.5	50.0		28.5			172
(5) Boys—pure science .....	48.7		51.3				226
Boys—applied science .....	39.7		60.3				229
(6) 16-yr.-old arts girls { Maint. Scs.	14.8	30.8	35.5	13.0	5.9		169
{ Ind. & D.G.	14.7	33.7	34.2	10.3	7.1		184
(7) Boys (arts) at London, Oxford and Cambridge .....	19.7	21.5	29.6	14.1	10.9	4.2	284
Boys (arts) at other universities .....	7.5	18.7	28.0	21.3	15.3	9.3	268
(8) Girls (arts) at London, Oxford and Cambridge .....	21.4	31.4	32.6	9.2	5.4		261
Girls (arts) at other universities .....	15.6	24.9	33.6	14.6	11.2		321
(9) Boys (science) at London, Oxford and Cambridge .....	9.0	16.8	27.3	19.0	18.8	9.0	399
Boys (science) at other universities .....	1.5	8.2	30.6	26.1	20.9	12.7	464

(Manuscript received, March 23rd, 1959.)



# A TYPE OF BIAS IN MARKING EXAMINATION SCRIPTS

By M. J. FARRELL AND N. GILBERT

**SUMMARY.** We consider the hypothesis that, in some circumstances, the variance of the marks awarded by an examiner will increase with the number of scripts he has marked. We test this hypothesis on the data of fifty-four different class-lists, and find a small but highly significant effect of the sort predicted. We consider various possible alternative explanations of the observed effect, and conclude that there is fairly strong evidence of the existence of such a bias.

## I.—INTRODUCTION.

IN this note we report some calculations made to test a hypothesis formed by one of us as to a likely type of bias in the marking of examination scripts. The hypothesis is that, in examinations which involve an element of judgment on the part of the examiner, the variance of the marks he awards will tend to increase with the number of scripts he has already marked. That is to say, the more scripts (of the relevant paper) he has marked, the more ready will he be to award extreme marks. We felt, on the basis of experience, that such an effect would be likely to arise through growing confidence on the part of the examiner. At first, even after a preliminary 'reading in,' the number of answers with which a given answer could be compared would be relatively small, and the examiner correspondingly cautious in the mark he gave; but as the number of answers available for comparison grew, so would the examiner's confidence. Of course, most examiners would do their best, by various devices, to counteract such a tendency but even so, some residual bias might persist. A similar effect might be produced by fatigue, but this would be more likely to be important in examinations like the General Certificate, where very large numbers of scripts have to be marked. In university examinations, we would expect the 'confidence' effect to be the important one. In any case, our test is for an increase in variance, however it may have been caused. If scripts are marked in alphabetical order, this would imply that extreme marks were more frequent in the later than in the earlier part of the alphabet. The same effect would appear, though less markedly, even if the scripts are not always marked in alphabetical order, so long as the earlier part of the alphabet tends, on the average, to be marked earlier.

## II.—METHOD AND RESULTS.

We sought evidence of such an effect in the class lists of the Cambridge Triposes. Each year a list is published of the successful candidates in each examination, the list being sub-divided according as the candidates obtained first, second or third class honours, or passed at a lower level. No list is published of those who failed outright, and this is a definite defect in the data from our point of view; but, fortunately, such cases are extremely rare, passing at a lower level being itself considered quite a serious failure. (Those who do not achieve third class honours are often sent down by their colleges.) We, therefore, used the number of those obtaining first-class honours, or passing at a lower level than third class honours as a measure of the number of extreme marks awarded.

Viewed purely as a psychological experiment, it would be much more efficient to use the actual marks awarded, as they contain a great deal more information than the class lists. We had no choice, as the marks themselves are not published, but in the present context there are two ways in which the class lists are actually preferable. First, many examiners would argue that the final classifications are considerably more accurate than the marks awarded, so that a test based on the former is fairer to the examiners. Secondly, since only the class lists are published, the bias will only matter in practice in so far as it is reflected in the class lists.

By classifying each successful candidate according to whether his class is extreme or central, and according to whether the initial of his surname comes before L or after K in the alphabet, we can summarise the information in each class list in a two-by-two table like that of Table 1. From each such table we calculate the statistic  $x$ , where

$$x = (ad - bc) \sqrt{\frac{a+b+c+d}{(a+b)(c+d)(a+c)(b+d)}}$$

The values of  $x$  for each of the seventy-two class lists considered are given in Table 2.

TABLE 1

Class \ Initial	A to K	L to Z
Extreme .....	a	b
Central .....	c	d

The total number of candidates on each class list was usually between 100 and 200, except in Economics Part II (about 70) and Natural Sciences (300 to 500). In the case of the Economics Tripos, the frequency of extreme classes was too low for the  $\chi^2$  approximation, on which our main significance test depends, to hold, so we included third class honours in the 'extreme' group. This is legitimate since the definition of 'extreme' is in any case arbitrary; the proportion of seconds and thirds is much greater in Economics than in other Triposes; and so far as one can see by inspection, it does not seriously alter the general picture.

The problem of combining the information from these seventy-two two-by-two tables so as to obtain an efficient test of our hypothesis is an interesting one. The frequencies cannot, of course, be added together to yield one large two-by-two table, since the probability of extreme classes certainly varies between faculties and may well do so over time. The most commonly used method is to observe that, since each  $x^2$  is distributed approximately like  $\chi^2$  with one degree of freedom, if we can assume the  $x$ 's to be independent, then  $\Sigma x^2$  is distributed like  $\chi^2$  with seventy-two degrees of freedom. This gives a test, but an exceedingly inefficient one: for it merely considers the size of the deviations of  $x$  from its expected zero value, and ignores their sign. It is thus very insensitive to small, persistent deviations of  $x$  all in the same direction, which is the sort of effect we have here.

TABLE 2  
EQUIVALENT NORMAL DEVIATES DERIVED FROM 2X2 CONTINGENCY TABLES.

	1948	1949	1950	1951	1952	1953	1954	1955	1956	Total
Mechanical Science Tripos Preliminary Examination	+0.193 +0.795	-1.297 +1.604	-1.254 -0.994	+0.644 -1.584	-1.542 +0.811	-0.573 -0.492	-1.100 -0.063	-2.038 -1.177	-0.171 -2.122	-7.138 -3.222
Economics Tripos .....	+0.509 -0.828	-0.388 -0.302	+0.212 +0.022	-1.429 -1.893	-1.654 +0.135	+0.931 -0.590	-2.166 -0.039	+0.956 -0.197	-2.542 +0.988	-5.571 -2.704
Mathematics Tripos .....	-1.561 +0.437	+0.142 +0.691	-0.653 +0.022	-0.836 +0.124	-0.582 +1.326	-0.695 -1.528	+0.479 +1.789	-0.091 -0.646	-2.056 -0.289	-5.853 +1.926
Natural Sciences Tripos ..	-0.194 -0.176	-1.640 -1.324	+0.128 +0.647	-0.866 -0.443	-0.092 -0.826	-0.301 -2.270	+1.585 -0.095	+0.502 -0.496	+1.459 -0.292	+0.581 -5.275
Total .....	-0.825	-2.514	-1.870	-6.283	-2.424	-5.518	+0.390	-3.187	-5.025	-27.256



Just how inefficient the 'adding  $\chi^2$ ' test can be is illustrated in the present case by comparing it with a simple combinatorial test.  $\Sigma x^2 = 83.62$ , and the chances of obtaining a value at least as large by chance are better than 1 in 7. However, if the deviations were purely random, the number of positive values of  $x$  in a sample of seventy-two would be binomially distributed with  $p = \frac{1}{2}$ ,  $n = 72$ , and the chance that as few as twenty-five of the  $x$ 's should be positive would be less than 1 in 100. Thus, the rough-and-ready combinatorial method, which ignores the size of the deviations, is yet much more sensitive than the 'adding  $\chi^2$ ' method.

The method actually used, due to Yates (1955), successfully takes account of both sign and magnitude. It turns on the fact that if  $x^2$  is distributed like  $\chi^2$  with one degree of freedom, then  $x$  is normally distributed with zero mean and unit variance. The values of  $x$  in Table 2 do, in fact, show a normal distribution with unit variance, although the mean is not zero. Now in the absence of bias,  $\Sigma x$  should be normal with zero mean and variance 72, so that  $\Sigma x (= -27.256)$  shows a deviation from its expected value of more than three times its standard deviation. In fact, the probability of its chance occurrence is about 1 in 1,000. The analysis of variance in Table 3 shows that there is no significant heterogeneity—the same bias seems to persist in all the examinations and in all the years considered. The observed residual mean square may be compared with its theoretical value of unity.

TABLE 3

Analysis of Variance	d.f.	S.S.	M.S.
Mean .....	1	10.32	10.32
Between years .....	8	4.91	0.61
Between faculties .....	3	1.53	0.51
Between examinations, within faculties .....	4	6.57	1.64
Residual .....	56	60.29	1.08

The probabilities quoted are those appropriate to testing for a change in variance, using a 'two-tailed' significance test. Now, in fact, our hypothesis was that there would be an increase in variance; it discounted the possibility of a reduction in variance. Consequently we should, strictly, apply a 'one-tailed' significance test, obtaining a chance for the observed increase in variance of 1 in 2,000. The question is in any case of purely academic importance, since a probability of 1 in 1,000 is small enough to convince most people. Most human minds seem to jump from doubt to certainty at a significance level somewhere between 1 in 50 and 1 in 500.

Our tests assume that the values of  $x$  are all mutually independent. This is not strictly true, as many students take, in different years, both the examinations quoted for their faculty. However, the overlap is very incomplete, and in practice the  $x$ 's may be considered independent. The correlations (based on five or six degrees of freedom) between the appropriate values of  $x$  are as follows:

Mechanical Sciences Tripos .....	+0.04
Economics Tripos .....	-0.26
Mathematics Tripos .....	+0.05
Natural Sciences Tripos .....	-0.25

The possibility must be considered that our results reflect, not an increase in the variance but a shift in the mean of the marks awarded. When the number of firsts exceeds the number of 'passes at a lower level,' a general increase in the marks awarded to L-Z candidates (as opposed to the A-K's) will result in a corresponding increase in the extreme classes at the expense of the seconds and thirds. That is, a shift in the average severity of marking will give an apparent change in the variance, simply because of an inequality in the frequencies of the two extreme classes.

It is possible to test this by constructing another sort of two-by-two table for each class list, classifying classes as 'good' (first or second) or 'bad.' The statistic  $y$  can be calculated from each such table analogously to  $x$ , and from the values of  $y$  the hypothesis of a mean shift can be tested. A little care is needed. The most favourable hypothesis (although rather an unlikely one *a priori*) is that those faculties where the upper extreme is more common than the lower display a rise in the mean, and the others a fall. Reversing the sign of  $y$  (wherever necessary) in accordance with this hypothesis, the most effective test of this rival hypothesis is to correlate  $x$  and  $y$ . As  $x$  and  $y$  have the same sign in only thirty-seven of the seventy-two cases, the possibility can be rejected. (These arguments hold so long as the distribution of marks is unimodal and moderately symmetrical).

### III.—DISCUSSION.

We may conclude that there is definite evidence that the variance of the marks awarded to the L to Z group is greater than that for the A to K's, but in the absence of any control experiment, it is impossible to rule out the possibility that this is a property of the candidates rather than the marking. That is, it is logically possible that those whose surnames come late in the alphabet shew a greater variation in aptitude for the subjects in question.

One way in which such a situation might conceivably arise is that some particular national or racial group might (a) have a substantially different distribution of surnames between the two halves of the alphabet, and (b) a substantially different dispersion in the aptitudes of its candidates, as compared with the rest of the candidates. If both these differences were large enough, if the national or racial group in question formed a sufficiently large proportion of all candidates, and if the differences were in the right direction, this could produce the observed difference.

For example, if candidates from overseas contained a larger proportion with surnames in the latter part of the alphabet than candidates from the United Kingdom, and if they shewed a greater variation in aptitude, then this would produce an effect of the kind observed. But, bearing in mind the relatively small number of overseas candidates, and the (at least) moderate similarity of the distribution of surnames between the parts of the alphabet, the dispersion of aptitudes would have to be spectacularly greater to account for even the small effect we have observed.

It is obvious that a large number of such hypotheses can be constructed—in themselves more or less plausible, but unlikely to be of sufficient magnitude to explain the observed effect. It would be very difficult, without much more information than is available, to test any one such theory; and, it would, of course, be impossible to test exhaustively all the conceivable hypotheses of this type. The reader must decide for himself how much weight to allow to such possibilities.

It is perhaps of interest to see how large the bias is. As we have said, the simple adding together of frequencies from different class lists is not strictly defensible; but it may serve to give an impression of the rough order of magnitude of the effect. In all years, 20.1 per cent. of the A to K group obtained extreme classes in Part I of the Mechanical Science Tripos, compared with 20.8 per cent. of the L to Z group. This is indeed a small bias, but its very smallness serves to emphasize how persistent it must be to achieve such high levels of statistical significance.

The bias that is so small is, of course, that which shows itself in the class lists—the important quantity from a practical viewpoint. But from a purely psychological point of view the bias in the original marking is more interesting, and this may well be much larger. For the proportion of the latter that shews in the class lists will be reduced by any attempts by the examiners to avoid bias, for example, by re-reading, and in particular by any departure from alphabetical order. Thus, if a fraction  $f$  ( $< \frac{1}{2}$ ) of the examiners read their scripts in counter-alphabetical order the bias is reduced by (approximately) a factor  $1-2f$ . (This suggests what is probably the simplest practical remedy—to have half the examiners read the scripts in counter-alphabetical order and the other half in alphabetical order. The bias in the aggregate mark is then reduced to a second order quantity depending on differential bias between examiners of papers and differential performance between papers.) It should also be noted that the subjects chosen are all more or less scientific ones, where the element of judgment in marking is relatively small. In more literary subjects one might expect a larger bias, although this might be offset by the examiners' greater awareness of its possibility.

#### IV.—REFERENCES.

- CAMBRIDGE UNIVERSITY REPORTER (1948-1956). Cambridge Univ. Press.  
YATES, F. (1955). A note on the application of the combination of probabilities test to a set of  $2 \times 2$  tables. *Biometrika*, 42, 404.

(*Manuscript received, April 6th, 1959.*)



# ZERO ERROR IN MORAY HOUSE VERBAL REASONING TESTS

By A. E. G. PILLINER

(*Department of Education, University of Edinburgh*)

J. SUTHERLAND

(*Moray House College of Education, Edinburgh*)

AND

E. G. TAYLOR

(*Formerly of the Department of Education, University of Edinburgh*)

**SUMMARY.** Non-equivalence in the performance of the samples of children used in standardising the several members of a series of similar tests results in lack of comparability between the standardised scores obtained with different tests in the series; all the standardised scores from one test may be positively or negatively biased relative to those from another. The bias is named 'zero error'.

The zero errors of twenty Moray House verbal reasoning tests, all relative to one taken as standard, are estimated and shown to be larger for more recent than for earlier tests. The increase is ascribed mainly to increased test sophistication in the later standardisation samples.

Local education authorities should take zero errors into account in their allocation procedures.

## I.—INTRODUCTION.

THE purpose of this article is to draw attention to a particular source of error which is likely to be present to some degree in all standardised scores or 'quotients' obtained when one or other of a series of standardised 'objective' tests, all bearing the same label, is administered. Examples of such series are the Moray House verbal reasoning tests or the National Foundation English attainment tests, and there are many others. The experimental part of this study is concerned only with the Moray House verbal reasoning tests, but there is little doubt that parallel studies with other series of tests would produce similar results.\*

Users of tests such as these are apt to assume rather lightly that the verbal reasoning (or whatever) performance level of two children is much the same if they obtain the same standardised score, the first child on test A in the series, the second on test B. Few nowadays would interpret this common standardised score as indicating *exactly* the same performance level for these children, since the concept of 'standard error of measurement' is a familiar one (see, for example, Guilford, 1950, p. 479). Many, however, are insufficiently aware of the possibility of a *bias* attaching to all the standardised scores obtained on one of the tests relative to those obtained on the other; that children scoring, say, 80, 100, and 120 on one test in the series might have scored, 84, 105, and 123, had a different test been chosen.

\* In a personal communication, Mr. D. A. Pidgeon, Senior Tests Officer of the National Foundation for Educational Research, tells us that the zero error studied here with Moray House verbal reasoning tests is observable with National Foundation tests also.

The occurrence of a bias in standardised scores obtained from one test relative to those from another in the same series may be misleading; for example, when a local authority uses one of a series of tests for the 'main' examination and another for the 'supplementary' examination for absentees; or when an allocation procedure involves a cut at some predetermined standardised score, regardless of the test used. Such procedures, too rigidly applied, are to be deprecated. They imply a faith in the consistency of calibration from test to test, which those who construct and standardise them are unable to share.

In the experimental investigation to be described, the mean standardisation bias of each of a number of Moray House verbal reasoning tests, relative to one chosen arbitrarily as standard, is evaluated. Since each bias reported is based on an experiment in which the subjects were a random sample from a larger group with a mean standard score of approximately 100 which corresponds to zero on the standard score scale; and since the bias under discussion is similar in some respects to the systematic errors sometimes occurring in physical measurement; we propose to call this standardisation mean bias 'zero' error.

## II.—ORIGIN OF ZERO ERROR.

The standardisation of a test initially involves its administration to a large sample of children (the nature of this sample will be discussed later), of the appropriate age range. This sample, called the 'standardisation' sample, is sub-divided by month of age, and within each month group the children's original or 'raw' scores are used to obtain their percentile ranks. These ranks are then transformed to a 'normal' distribution of standardised scores or 'quotients,' usually with a mean of 100 and a standard deviation of 15. After repeating this procedure for all month groups in the sample a conversion table or table of norms is prepared which relates the raw scores and ages of the children in this sample to the corresponding standardised scores. Similar conversion tables are prepared for each test in the series, using each time a different standardisation sample. Each table of norms thus prepared makes it possible to match the performance level of any child subsequently taking the corresponding test with that of some children in the standardisation sample. For example, if entering the table with some testee's age and raw score assigns him a standardised score of 100, his performance level on the test is equal to that of all the children in the standardisation sample whose several raw scores place them at the 50th percentile in their several month groups (Thomson 1932, Lawley 1950).

Suppose that on some particular occasion, a test was chosen, that it was used with a number of children, and that the mean of their standardised scores was 100. Now suppose that a different test in the series had been chosen and used on this one occasion with the same children and that their mean standardised score on this different test was 106. Note that, in fact, only one test has been administered and only one mean is available: *either* 100 *or* 106; and that the test user accepts whichever turns up in ignorance that the other *might* have done so. It is not inconceivable that some important decision may hinge on the result.

The obstacles in the way of achieving the ideal that any child tested should obtain the same standardised score, whichever test in the series happens to be the one administered to him are:

- (1) Differences in content from test to test, resulting in slightly different rank orders for the same children.
- (2) Non-equivalence in the standardisation samples.

With regard to (1), the imperfect correlation results in unbiased errors of measurement which can be estimated with some precision if the correlation is known. It is worth pointing out that it is, in fact, unknown; the ordinary test-retest correlation based on the administration of different tests to the same testees on different occasions under-estimates the correlation which should be employed, namely, that between the children's scores on the test used and the scores they would have obtained with a different test *on this one occasion*. An under-estimate of the correlation means an over-estimate of the standard error of measurement. In estimating the standard error of measurement we shall err, therefore, on the safe side by basing it on the usual test-retest correlation.

With regard to (2), by 'equivalence' would be meant that we can legitimately assume each standardisation sample to be representative of the same specifiable 'population' of children; a population against which it makes sense to measure the prospective testees in test performance. If obtainable, equivalent samples would be inter-changeable in respect of the tests, without affecting the tables of norms. Testees would be pitted virtually against the same population, whatever the test, and any discrepancies in their standardised scores if one test were used rather than another would be due entirely to differences in test content, as discussed in (1). With equivalent samples as here defined, there would be no systematic bias associated with all the standardised scores obtained with a particular test. 'Non-equivalence' would imply either that the standardisation samples are not representative samples from the same population or that they are samples from different populations. In either case the possibility would exist of systematic as well as unsystematic errors occurring.

The population to be represented by the standardisation samples consists of a large number of local authority year-groups, the mean standardised scores and the allocation procedures of which show remarkable diversity. A number of these authorities prefer to use tests already fully standardised and so must be excluded from consideration. In practice, therefore, it is impossible to achieve completely representative sampling and some approximate procedure must be used, such as the following. The use of newly constructed tests not yet fully standardised is restricted to those other authorities which are prepared to co-operate by making available the raw scores and ages of the children in year-groups working these new tests; in return these authorities receive specially constructed tables of norms based on the data supplied. From among the authorities in this sub-population which have chosen a particular new test a sample is selected which, on the basis of previous experience and current performance, is judged to be as nearly representative as possible of the general population of authorities. Finally, a 'national' conversion table for the new test is prepared from the pooled data of the authorities thus selected from the sub-population.

In spite of these difficulties, the zero error of any one Moray House verbal reasoning test relative to its neighbour in the list, is usually remarkably small; both, it should be noted, will have been standardised at about the same time. However, local authorities using new and old tests with consecutive year-groups (which may be expected to be about equal in mean performance), do experience a rise in mean standardised score from the new to the old test which indicates non-equivalence in the samples used for standardising these more widely separated tests. This non-equivalence results less from the selection of poor samples from a given population than from a slow change in the performance characteristics of the population itself, particularly since the early 1950's. It appears to be associated mainly with the observed wide-spread increase



during recent years in 'test-sophistication' which will be the subject of later discussion.

An experimental study designed specifically to estimate as precisely as possible the zero errors of commonly used tests was undertaken for the following reasons. First, such a study might contribute information useful in a general survey of test standards. Second, the results might be helpful to investigators using these tests for the study of population trends, earlier maturation, the effects of practice and coaching, and the like, including also the re-evaluation of the results of previous studies in these fields in which the zero errors of the tests used were not taken into account. Third, the publication of the figures obtained might not only contribute to a more general awareness among test users of the existence of zero error, but also give those responsible for allocation procedures very practical assistance in the interpretation of the results obtained.

### III.—THE EXPERIMENTAL INVESTIGATION.

The tests chosen were the verbal reasoning tests M.H.Ts. 33 and 39-58, for which 'national' norms had been prepared over the period 1945 to 1957. For six of these tests, namely M.H.Ts. 39-44, data were already available from Michael F. Moore's previous study (1955) of the practice effects resulting from the administration of six verbal reasoning tests in succession to the same group of children. Moore's study, while primarily concerned with practice effect, differed from previous studies of this kind in that his experimental design enabled him to estimate the zero errors of the tests used and to disentangle these errors from the practice effect. Each of six groups of children worked six tests in six different orders, the whole conforming to a latin square design, the reciprocal nature of which enabled Moore to isolate both practice effect and zero error simultaneously.\* The only draw-back to Moore's elegant design was that the final computations had to be restricted to the scores of six numerically equal groups of children all of whom completed all six tests. Illness during the administration period resulted in considerable loss. Profiting by Moore's experience, the present writers employed for the remaining tests a method which minimised wastage while retaining the advantages of the latin square.

The simplest way of carrying out the experiment would have been to use a number of large independent random samples, that is, equivalent samples. Sample 1 would have test A, sample 2 test B, and so on (Lindquist, 1953, p. 47). With proper randomisation, the expected mean performance level is the same for all samples, so that any observed differences among the standardised score means would be ascribable to zero error. This method was rejected because of the practical difficulties in the way of efficient randomisation and the lack of precision inherent in the design.

Instead, separate comparisons were carried out with two tests at a time. In each comparison, both test A and test B were administered to the same group of children using the 'cross-over' method. The children were divided at random into two numerically equal groups. Group I took first test A, then test B; Group II took first test B, and then test A. The difference between the overall mean score on A in its two positions and the corresponding overall mean for B gave the mean zero error, with practice effect eliminated.

Test A was always M.H.T. 40, chosen as standard. Test B was one of the other tests, the experiment being repeated, each time with different children, for all the tests except for Moore's six. The variable was standardised score ('quotient') on the 'national' norms in the test manuals.

\* The desirability of eliminating the effects of zero error in the test used for practice and coaching experiments was previously noted by Vernon (1954, p. 59).

The children and the tests were allocated to the groups as follows: Primary schools were chosen at random from the Edinburgh and Midlothian lists. Six of these schools, again chosen at random, were allotted to each comparison AB<sub>1</sub>, AB<sub>2</sub>, AB<sub>3</sub>, etc. Within each school the available children were allocated at random to two groups of the same size for each school, but not necessarily so from one school to another. Finally, within each school the group designation I or II, was decided by the toss of a coin. Group I's testing order was always AB and group II's BA.

The main advantages of this design are: first, the effect of absences is less catastrophic than in the larger latin square. Second, the replications in six schools for each pair of tests provides information about the 'interaction' between school and zero error, that is, the extent of variation of zero error from school to school, in addition to information about the 'main' effect, the mean zero error overall. Third, since a random half of the random sample of children engaged in the experiment took M.H.T. 40 first, valid inferences can be drawn about the performance on this test of the population sampled, and this information is useful in judging the validity of regarding the zero errors obtained in the experiment as generally applicable.

#### IV.—EXPERIMENTAL RESULTS.

Table 1 summarises the main results of the enquiry:

TABLE 1

ZERO ERRORS OF MORAY HOUSE VERBAL REASONING TESTS RELATIVE TO M.H.T. 40.

Standardisation Date	1945	1948	1948	1949	1949	1950	1950	1951	1951	1952	1952
N .....	186	132	132	132	132	132	132	162	142	162	156
M.H.T. ....	33	39 <sup>1</sup>	40 <sup>1</sup>	41 <sup>1</sup>	42 <sup>1</sup>	43 <sup>1</sup>	44 <sup>1</sup>	45	46	47	48
Zero Error ..	-0.17	-0.92	0	-0.41	0.28	-0.42	-0.14	-0.31	-0.85	-1.09	0.20
Standardisation Date	1953	1953	1954	1954	1955	1955	1955	1956	1957	1957	
N .....	170	122	192	186	174	182	182	210	168	188	
M.H.T. ....	49	50	51	52	53	54	55	56	57	58	
Zero Error ..	-4.38	-2.79	-4.82	-4.31	-5.76	-4.72	-6.70	-5.98	-6.11	-6.14	

N=number of children in each experiment.

<sup>1</sup> These tests were used by Mr. Moore, and the zero errors are those obtained by him. Permission to quote these results is gratefully acknowledged.

Mean standard error of zero errors relative to M.H.T. 40 =  $\pm .57$ . Mean standard error or zero errors derived as in (2) below = .80.

(1) A negative sign means that the mean standardised score on the test indicated is expected to be *less* by the zero error recorded than that obtained by the same group if M.H.T. 40 had been used instead.

(2) Zero errors relative to tests other than M.H.T. 40 may be readily obtained from Table 1. Thus, the zero error of M.H.T. 58, relative to M.H.T. 56 is  $-6.14 - (-5.98) = -0.16 \pm \text{S.E. } .80$ .

Table 2 presents a typical example of the fifteen analyses of variance.\* That in Table 2 results from the comparison of two tests, here M.H.T. 40 (test A) and M.H.T. 56 (test B). The analysis is the outcome of a design modified from the replicated  $2 \times 2$  latin square ('cross-over') design of Cochran and Cox (1950, pp. 112-116). The numbers of children in the six schools concerned were 32, 40, 40, 40, 36, 30, and 32; within each school the children were allocated to groups I and II, in equal numbers. Group I's test order was AB; group II's, BA.

TABLE 2

ANALYSIS OF VARIANCE OF STANDARDISED SCORES OBTAINED BY CHILDREN IN SIX SCHOOLS ON M.H.T. 40 AND M.H.T. 56.

Source	d.f.	S.S.	M.S.	F.
Schools .....	5	18,256.7	3,651.3	75.0(1,5)** 4.13(5,198)**
Groups within schools .....	6	619.2	103.2	
1st vs. 2nd occasion .....	1	367.7	367.7	
Schools $\times$ occasions .....	5	119.9	24.0	
Tests .....	1	3,750.1	3,750.1	
Schools $\times$ Tests (S $\times$ T) .....	5	249.8	50.0	
Child Totals within Groups within Schools	198	37,573.7	189.8	
Child scores $\times$ Tests (C $\times$ T) within Groups within Schools .....	198	2,397.9	12.1	
Total .....	419	63,335.0	—	

The F's recorded indicate that the overall mean zero error is significant and that the zero error varies significantly among schools.

In Table 3 are reported the standardised score means for the six schools used for the M.H.T. 40—M.H.T. 56 comparison. Each school mean is based on the sum of all scores obtained by the children in the school, half of the scores being in AB order, the other half in BA order. Table 3 gives also the mean zero error for each of the six schools. The regression of zero error on school mean should be noted.

TABLE 3  
SCHOOL MEANS AND ZERO ERRORS.

School	1	2	3	4	5	6	All
N .....	32	40	40	36	30	32	210
Mean .....	110.23	108.19	108.14	100.28	96.07	92.38	102.99
Zero Error .....	-7.78	-6.53	-7.17	-6.28	-3.27	-4.19	-5.98

\* The scores from M.H.T. 39-44 were analysed by Moore in a single table.

#### V.—DISCUSSION AND IMPLICATION OF RESULTS.

Each of the zero errors in Table 1 is based on a comparison between the arbitrary standard M.H.T. 40 and one other test administered to the same sample of children. The care taken over sampling procedure makes it reasonable to



accept these observed zero errors as estimates of the population values, the population consisting of the pool of two year-groups of children of appropriate age. A sample of 1,291 children drawn from these two year-groups—those who took M.H.T. 40 first—had a mean on this test of 105.5, which we may take as a good estimate of the population mean on M.H.T. 40. Had this population of two year-groups taken one of the later tests with a zero error of 5.6 points, the expected mean would have been close to 100, this expectation tallying with our knowledge from other sources of the mean performance levels on later tests of these year-groups. The weight of the statistical evidence, namely: the adjusted mean of about 100; the observed standard deviation of 13.3; the coefficients of skewness ( $\beta_1 = -0.29$ ) and kurtosis ( $\beta_2 = 2.52$ ); all this makes it probable that the Table 1 zero errors are applicable to most current year-groups.

These zero errors are themselves subject to error of measurement. Omitting Moore's six tests, there remain fifteen comparisons with M.H.T. 40 in most of which the Schools  $\times$  Tests interactions were significant (the mean square of 50.0 in Table 2 is rather larger than average). The significance of these interactions has important consequences for the precision of the zero error estimates, since it implies a significant heterogeneity of zero error from school to school in each comparison which must be inferred in the population sampled; the appropriate errors of measurement must, therefore, be based on the mean square for the Schools  $\times$  Tests interactions. Thus, in the M.H.T. 40—M.H.T. 56 comparison, the estimated zero error of 5.98 has a standard error of  $\sqrt{(2 \times 50/210)} = .690$ . The average of all such standard errors is 0.570, which is appropriate for zero errors as recorded in Table 1 (i.e., relative to M.H.T. 40). For zero errors obtained by taking the difference between two recorded errors (i.e., relative to a test other than M.H.T. 40), the mean standard error is  $.570 \sqrt{2} = .80$  approximately.

A study of the regression of zero error on school mean in the fifteen comparisons reveals in some, but not all, a tendency for zero error to increase with school mean. The data of Table 3 provide a striking example; the regression is highly significant. In other comparisons the evidence is less clear, but the general trend is in the same direction. Fortunately, the observed regressions are insufficiently large to necessitate adjustment of observed zero error over the range of local authority mean quotients likely to be met with.

Table 2 illustrates a further point of some importance. The final mean square in the table (12.1) measures the mean error variance of measurement of children's individual standardised scores. This error variance implies a correlation between M.H.T. 40 and M.H.T. 56 scores of .95, referred to a population standard deviation of 15.\* Correlations, similarly estimated, between M.H.T. 40 and each of the other tests, range from .92 to .96 with average .94. This uniformity of correlation is evidence that all the tests will place the same set of testees in much the same rank order as M.H.T. 40, and hence as each other. For the validities of the tests in the series with respect to a common criterion, this uniformity is important. There is no evidence of that cumulative error, producing a 'drift' from the criterion, against which Heim (1954, p. 109) rightly warns us.

The zero errors in Table 1 show the effect on the mean standardised score of the same group of testees if one test, rather than another, happens to be chosen. For example, the mean on M.H.T. 58 is estimated as 6.14 ( $\pm$  SE .57) points lower than that on M.H.T. 40; and, similarly, the mean on M.H.T. 54

\*  $r = (15^2 - 12.1)/15^2 = .95$ . This follows from: error variance =  $s^2(1-r)$ .

as 3.63 ( $\pm$ SE .80) points lower than that on M.H.T. 47. These apparent differences in 'difficulty' indicate a rise in the performance level of the samples available for standardisation from the earlier to the later tests. The same group of testees, pitted against a later instead of an earlier standardisation sample, will do relatively less well and so obtain lower standardised scores on the later test. There are indications in Table I of a very slight upward trend in the zero errors over the first eleven tests listed, and a quite sharp discontinuity between that of M.H.T. 48 and that of M.H.T. 49. The upward trend then persists, but appears to be levelling out over the last four tests in the list.

There can be little doubt that this rise in performance level reflects mainly the effect of test-sophistication, increasing from earlier to later standardisation samples. Children in later samples have become more familiar than their predecessors with the type of material included in tests such as these through more extensive coaching or practice on similar material. It is significant that the discontinuity in zero error between M.H.T. 48 and M.H.T. 49 occurred at a time (1953) when many local authorities (including those used in the standardisation samples) were introducing full-scale practice tests as a preliminary to their allocation procedures; the time of the public controversy over coaching and practice (Vernon, 1952). The further upward trend and the flattening towards the end of the series suggests that saturation may now have been reached, though we cannot be certain of this.

We can only speculate about other influences contributing to the rise in performance level reflected in increasing zero errors. Tanner (1958, p. 1689) has pointed out that as a result of better nutrition and perhaps because the world is getting warmer, children in Western Europe and America are maturing physically at an earlier age than their predecessors; according to him the presumption is that there is a similar trend in mental development. If this is so, the effect on performance and hence on zero error will reinforce that of increasing test-sophistication.

We turn now to the practical uses to which a knowledge of these zero errors can be put. In the first place they should be useful to those responsible for testing groups of children when comparisons have to be made between two groups of children tested with different tests. A knowledge of the zero error of one test relative to the other will enable the tester to adjust one of the observed group means so that the comparison between them is unbiased. The step from group means to individual scores is a little hazardous; this study is concerned with biases in mean scores only, and provides no evidence that the bias is constant throughout the range of individual scores; indeed, we present some slight evidence that zero error, like practice effect, is positively correlated with score. However, if the step from mean scores to individual scores is made, it will be in the right direction even if its magnitude is not entirely correct. Thus, in the case of a child who moved from one area where he obtained a standardised score on M.H.T. 47 of 114 to another where the test used was M.H.T. 57, it would be fairer to adjust his standardised score to 119, than to leave it at 114 in placing him in relation to the children in his new year-group. Again, in the case previously mentioned, of a supplementary examination administered to absentees from the main examination, it would not be unreasonable to diminish the score of each child taking the supplementary examination by 4 points if the supplementary test were M.H.T. 50 and the main test M.H.T. 55.

Zero error may need to be taken into account in interpreting the results of research investigations; for example, the study by Wiseman and Wrigley (1953) of the effects of coaching and practice based on test booklets containing

material alleged to be similar to that incorporated in verbal reasoning tests. They used a number of Moray House tests to evaluate the practice and coaching effects at various stages in their experiment, and had reason to suspect, from the general pattern of group mean scores, standardisation errors in M.H.T. 39 and M.H.T. 40. If we adjust (where we have the data to do so) the mean scores obtained by Wiseman and Wrigley in the test run for the children who had simple practice we obtain the following :

	Initial Test	Practice Tests						Final Tests
M.H.T. . .	43	37	38	39	40	41	42	44
W. and W's Means . . .	99.62	104.7	108.88	105.73	110.84	109.05	110.57	110.71
Zero Errors	-0.42	*	*	-0.92	0	-1.41	0.28	-0.14
Adjusted Means . . .	100.04	(104.07)	(108.88)	106.65	110.84	110.46	110.29	110.85

The adjustments have had the effect of bringing the M.H.T. 39 mean more into line with the general trend, and the flattening out from M.H.T. 40 is slightly more regular than with the unadjusted array of means. The authors' main conclusions remain unaltered, but it is easy to see how misleading the results might have been had tests with widely discrepant zero errors been chosen for the experiment.

Finally, the results of the present study have implications for the problem of test standards generally. The upward trend of zero errors is evidence that the level of performance on tests of this kind has risen generally and may continue to rise. Test standardisations must of necessity be based on the performances of contemporary samples of children ; and these, it is argued here, are not equivalent to earlier samples. Are those responsible for the standardisations to continue to assign to each new test an arbitrary mean standardised score of 100 in spite of this known non-equivalence ? Or are they to relate the test means to that of some previous test, standardised on a sample of children of an earlier ' vintage,' assigning to that test only—as in this study—an arbitrary mean of 100 ? To adopt the first course is to fail to take account of and, indeed, to obscure the undoubted rise in performance level the country over ; to adopt the second is to accept the necessity of embarking on an extension of the work reported here, to co-relate all present and future tests in this and all other series of tests from whatever source—a formidable task. The second course, though perhaps preferable, presents difficulties, particularly with regard to tests of attainment. Test constructors are continually trying to improve their tests, to incorporate in them the results of recent research findings, to increase their validity, to make them more liberal in conception and to minimise undesirable ' backwash ' effects on school curricula. To relate performance on some future English attainment test, say, to that on an early member of the series, when the tests have little in common except their name, would seem unrealistic.

The time has come for urgent consideration by psychometrists generally of issues such as these.



## VI.—REFERENCES.

- COCHRAN, W. G., and COX, G. M. (1950). *Experimental Designs*. New York : Wiley.
- GUILFORD, J. P. (1950). *Fundamental Statistics in Psychology and Education*. New York : McGraw-Hill.
- HEIM, A. W. (1954). *The Appraisal of Intelligence*. London : Methuen.
- LAWLEY, D. N. (1950). A method of standardising group tests. *Brit. J. Educ. Psychol.*, 2, 86-89.
- LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education*. Boston : Houghton Mifflin.
- MOORE, M. F. (1955). *The Effect of Coaching and Practice on Moray House Intelligence Test Scores*. B.Ed. Thesis, Edinburgh University Library.
- TANNER, J. M. (1958). *Times Educational Supplement*, November 21st, 1958.
- THOMSON, G. H. (1932). The standardisation of group tests and the scatter of intelligence quotients. *Brit. J. Educ. Psychol.*, 2, 92-112, 123-138.
- VERNON, P. E. (1952). *Intelligence Testing*. *Times Educational Supplement* reprint of articles and letters on practice and coaching.
- VERNON, P. E. (1954). Symposium on the effects of coaching and practice in intelligence tests. *Brit. J. Educ. Psychol.*, 24, 57-63.
- WISEMAN, S., and WRIGLEY, J. (1953). The comparative effects of coaching and practice on the results of verbal intelligence tests. *Brit. J. Psychol.*, 44, 83-94.

(Manuscript received, April 10th, 1959.)

# DIFFERENCES IN ATTAINMENT BETWEEN PRIMARY-SCHOOLS IN MIXED-LANGUAGE AREAS: THEIR DEPENDENCE ON INTELLIGENCE AND LINGUISTIC BACKGROUND

By D. G. LEWIS

(*Department of Education, Queen's University, Belfast.*)

**SUMMARY.** Differences in attainment in English and arithmetic between primary schools in mixed-language areas of Wales are studied with respect to differences in both intelligence and linguistic background. School attainments tended to increase with intelligence, though not regularly; and the school differences in intelligence did not account fully for the differences in attainment. Generally, the attainment of schools with pupils of a strongly Welsh background—i.e., bilingual pupils—was lower than that of other schools, this tendency being strong in English and slight in arithmetic. It is concluded that a bilingual environment may be regarded as an important factor with respect to primary-school differences of attainment in English which cannot be accounted for by intelligence.

## I.—INTRODUCTION.

THIS investigation is based on the results of standardized tests of English, arithmetic and (non-verbal) intelligence which were given to all the 10-year-old pupils from sixteen primary schools in the counties of Cardiganshire, Carmarthenshire, Merionethshire, and Pembrokeshire.\* Linguistic background was assessed by means of a language questionnaire, and test scores from a total of 375 pupils were obtained. The schools differed considerably with respect to the linguistic background of their pupils, some drawing most of their pupils from Welsh-speaking homes, others having only a small minority of such pupils. None of the schools were 'Welsh' schools, where the medium of instruction is Welsh, and all the pupils could understand and speak English with reasonable fluency. Pupils with a strongly-Welsh background are thus bilingual, while those with a background of limited 'Welshness' are monoglot. The aim of this investigation is to see how any differences in attainment between the schools relate to corresponding differences in intelligence, and, in particular, to see whether the attainment differences are in excess of what might be expected from differences in intelligence. If this is so, then changes in the linguistic background of the pupils (i.e., the varying degrees of 'Welshness' in their social environment), may appear as an important factor. Certainly, this possibility should be considered before attributing any unexplained attainment differences entirely to the teaching factor.

## II.—DESCRIPTION OF TESTS<sup>1</sup> AND QUESTIONNAIRE.

### A.—*The English and Arithmetic Tests.*

Two tests of attainment in English were given:

- (i) *Schonell's Silent Reading Test B (Test R4)*, which provides a measure of

\* The data were made available by the Department of Education, University College of Wales, Aberystwyth. The writer wishes to thank the department and, in particular, Mr. J. R. Morrison and Dr. Hywela Saer (lecturers in the department) for much helpful discussion during the course of the investigation.

attainment in the comprehension of English. It consists of twenty items, of which the testee is required to select, from a number of alternatives, words which make the best sense when put in blank spaces in a short passage of prose. The test has a time limit of 15 minutes. It will be referred to as English Comprehension.

(ii) *Moray House English Test 21 (M.H.E.21)*, which provides a measure of attainment in the general usage of English. It is a test designed especially for grammar-school selection and, during recent years, has been in private use amongst education authorities for this purpose. It consists of 120 items (synonyms, spelling, punctuation, vocabulary, idioms, etc.) and has a time limit of 40 minutes. It will be referred to as English Usage.

Two tests of attainment in arithmetic were given :

(i) *Schonell's Essential Mechanical Arithmetic Test (Form A)*, which provides a measure of attainment in the basic arithmetical processes.

(ii) *Schonell's Essential Problem Arithmetic Test (Form A)*, which provides a measure of attainment in the understanding and solving of written arithmetical problems.

Both these tests consist of fifty items and have a time limit of 30 minutes. They will be referred to as Mechanical Arithmetic and Problem Arithmetic respectively.\*

#### B.—*The Intelligence Test.*

Jenkins' Scale of Non-verbal Ability (N.F.E.R., 1947) was used as a measure of intelligence. Previous investigations have suggested that verbal tests of intelligence are not suitable for measuring the intelligence of bilingual children—or at any rate for assessing their capacities relative to those of monoglot English-speaking children (see, for example, Saer (1922 and 1923), Smith (1923), Barke (1933), Barke and Parry-Williams (1938), and Jones (1952)). Jenkins' test is non-verbal and diagrammatic in character, the testees having to select the correct figure for each item. It consists of eighty-five items, arranged in five sub-tests separately timed, the total time required being 30 minutes. An alternative version of the test, in which all instructions were given in Welsh, was available for all children who preferred to use it.

#### C.—*The Questionnaire.*

A language questionnaire constructed along lines similar to Hoffmann's Bilingual Schedule (Hoffmann, 1934) was administered to assess the (English-Welsh) linguistic background of the children tested. The questionnaire had previously been used in Caernarvonshire to select a Welsh-speaking group for the purpose of standardizing certain tests (see Miles, 1953). From the questionnaire scores children whose background is strongly Welsh can be distinguished from those whose background is mainly English or linguistically mixed. A Welsh version of the questionnaire was provided for all who preferred to answer it in Welsh.

### III.—ANALYSIS OF TEST RESULTS.

School attainments were found to differ considerably for each of the four attainment tests.† There was a tendency for the school attainments to increase

\* See Schonell (1950), 71-77 and 87-95, for further details of these tests.

† In one case, however, (English Comprehension) the differences just failed to attain significance at the 5 per cent. level ( $F=1.60$ , for 15 and 359 degrees of freedom).



with intelligence. In no case, however, was this increase a regular one. The correlation of each set of school attainment means with the corresponding intelligence means was as shown in the following table.\*

TABLE 1

	Compre- hension	Usage	Mech. Arith.	Problem Arith.
Correlation Coefficient (between attainment means and intelligence means)	.58	.60	.18	.65

Differences in attainment between schools may clearly be considered as being dependent, to some extent, on the corresponding differences in intelligence. It seemed desirable, therefore, to adjust the attainment differences to make allowances for the school differences in intelligence. This was done by the analysis of covariance (see, for example, Lindquist, 1940, 180-207). The school attainment means were thus compared in relation to their intelligence means and the within-schools regression of attainment on intelligence.† In the case of the English tests this comparison is illustrated by Figures I and II. It is seen that, in both tests, the dispersion of attainment means around the regression line is considerable. The dispersion, in fact, was found to be statistically

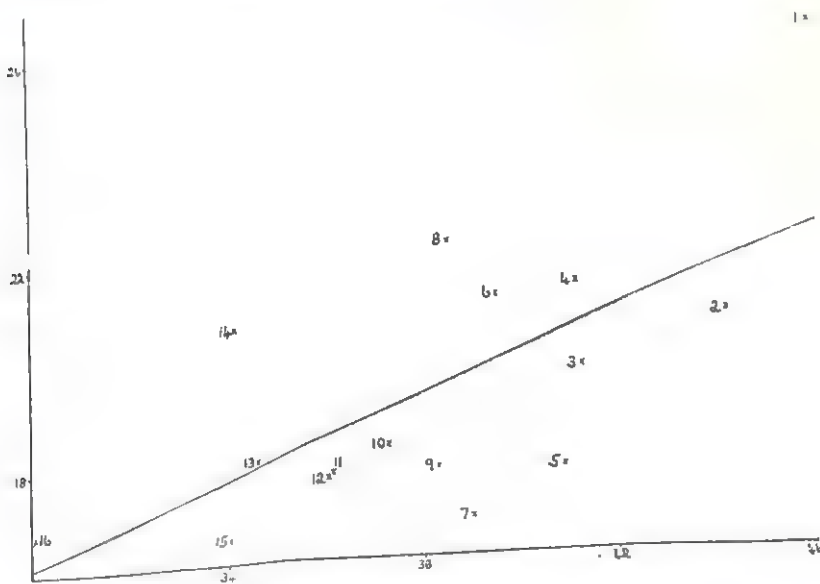


FIG. I. English Comprehension and Intelligence

\* In these correlations each mean was weighted by the number of pupils tested in that school.

† By this is meant an average regression of all the separate regressions of attainment on intelligence within each of the schools. It is essentially an estimate of what the regression would be if all the pupils had been taught in the same school (see Lindquist, 1940, 186).

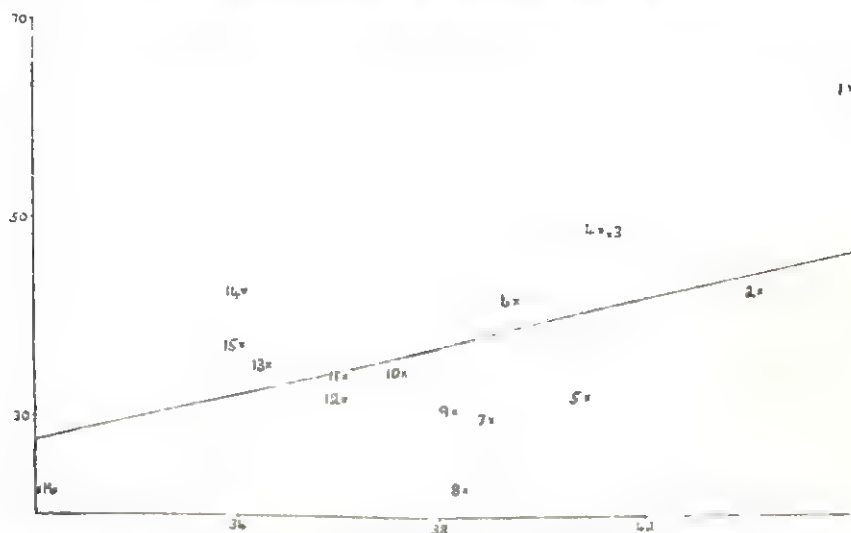


FIG. II. English Usage and Intelligence.

significant (at the 1 per cent. level) for all four attainment tests. Differences in intelligence between schools do not account fully for their differences in attainment.

In both the English tests schools 1 and 14 are seen to have done markedly better than might be expected on the basis of intelligence alone, the points representing these schools being well above the regression line in both diagrams. At the other extreme the points representing schools 7, 9 and 5, are seen to be appreciably below the regression line in both diagrams. Schools 4 and 3 may be noted as having a high adjusted mean\* on Usage. School 8 yields the curious result of having done well on Comprehension, being third in the order of adjusted means and yet very poorly on Usage. All other means on the two English tests seem to be fairly close to the regression lines.

In both the arithmetic tests schools 1 and 9 were found to have done considerably better than their intelligence expectation, i.e., in diagrams corresponding to those shown for the English tests the points representing these schools would appear well above the regression line. At the other extreme the attainments of schools 7 and 2 were found to be markedly less than that expected on the basis of intelligence. Again, school 14 had a high adjusted mean on Mechanical Arithmetic and school 6 a low adjusted mean on Problem Arithmetic. All other means in the arithmetic tests were fairly close to the regression lines.

Thus, for the results as a whole, school 1 has done extremely well, its adjusted mean being actually the highest in all four tests. Although this school has the most intelligent group of 11-year-olds (its intelligence mean is the highest as may be seen from the position of the point representing this school on the extreme of the figures) so that high attainment means are to be expected, nevertheless, the actual attainment means are higher still. School 7 on the other hand, can be noted as having attainment means appreciably below what might be expected from intelligence. The results of the linguistic-background questionnaire can thus be considered to see whether they afford a partial explanation of these school differences.

\* i.e., high in relation to the regression line, which provides the basis of the adjustment for school differences in intelligence.

## IV.—THE QUESTIONNAIRE RESULTS.

Considerable differences in linguistic background would naturally be expected within each school, but a convenient index of the linguistic background of each school group as a whole is provided by the percentage of the group whose background, as assessed by the questionnaire, is strongly Welsh. In the report of the administration of the questionnaire in Caernarvonshire all who scored from +51 per cent. to +100 per cent. were grouped together 'definitely Welsh' (Miles, *ibid.*, 20). However, the writer, in a recent paper,\* has put forward strong reasons for favouring a more narrowly-defined and homogeneous 'Welsh' group (+90 per cent. to +100 per cent. on the questionnaire). Consequently, the percentage of pupils for each school in this 'Welsh' group was taken as a *linguistic index* of the school.† These linguistic indices—which can be regarded as measures of the 'Welshness' of background of the groups tested—are shown, together with their ranks, in the following table:

TABLE 2

School No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Linguistic Background Index	20.0	14.6	15.1	12.2	36.2	25.0	41.6	36.2	43.7	50.0	31.2	24.2	18.7	9.8	57.1	26.7
Index-rank	11	14	13	15	5	9	4	5	3	2	7	10	12	16	1	8

It is seen that schools 1 and 14, already noted as having high adjusted attainment means on both the English tests, have low index-ranks 11 and 16 respectively. Again, schools 7, 9 and 5, noted as having low adjusted means on both the English tests, all have much higher index-ranks of 4, 3 and 5 respectively. It is clear then that schools which have done markedly well on the English tests, i.e., noticeably better than would be expected on the basis of their pupils' intelligence alone, are schools with a low index-rank and, again, schools whose attainments in English are noticeably below their intelligence expectations are schools with a high index-rank.‡ In the same way, schools 4 and 3, which were noted as having high adjusted means on Usage have low index-ranks 15 and 13, respectively, while school 8, which has a surprisingly low attainment on this test, has a fairly high index-rank of 5. Possibly, too, the difference between this school's attainment on the two English tests can be related to its linguistic-background index. If the usage of English is more dependent on the home than is the comprehension of the language (which may be the more dependent on the school in strongly Welsh-speaking areas), then schools with high index-ranks may reasonably be expected to show attainment differences of this kind. At any rate school 8's shift in position from Comprehension to Usage may be contrasted with that of schools of low index-ranks (e.g. 3, 4 and 14) which show no such shift.

\* This *Journal*, XXIX, 17-22.

† More strictly, of the ten-year-old pupils of the school.

‡ On the other hand, the fact that a school has a high (or low) index value does not necessarily imply that its attainment in English is below (or above) its intelligence expectation. Thus, school 15, with an index-rank of 1, has an attainment mean on Usage slightly above its intelligence expectation.



In the arithmetic tests the tendency for high adjusted attainment means to be associated with low index-ranks and low adjusted attainment means with high index-ranks is not evident from inspection. Thus, while school 1, with an index-rank of 11, is at the top end of the attainment scale, so also is school 9, with an index rank of 3. Again, at the other extreme, schools 7 and 2, both with low adjusted means on both tests, have index-ranks of 4 and 14 respectively. No association between adjusted attainment means and the index-ranks thus suggests itself.

However, the relative magnitude of these tendencies for all four tests can be assessed by correlating the adjusted attainment means with linguistic background. Spearman's rank correlation, obtained by correlating the ranks of the adjusted means with the ranks of the linguistic-background indices, is shown below for each test. The standard error, calculated from the formula involving the equivalent Pearson  $r$ , is also shown (see Guilford, 1936, 341).

TABLE 3

	Comprehension	Usage	Mech. Arith.	Prob. Arith.
$\rho$	— .51	— .55	— .12	— .05
S.E. of $\rho$ . . . . .	.19	.18	.26	.26

The first two of the above correlation coefficients are, numerically, considerably larger than their standard errors and may be regarded as significantly negative. This is obviously not so for the last two. This confirms that the tendency for high adjusted means go with low index-ranks and for low adjusted means with high index-ranks is a marked one in the case of each of the English tests, but that it is only a slight one in the case of the arithmetic tests.

One can conclude then that linguistic background can be regarded as an important factor of the school differences in attainment in English which are not explained by differences in intelligence. With respect to the corresponding attainment differences in arithmetic linguistic background is of less importance. There is some evidence of a tendency similar to that operating in English, but further research, with an analysis of results from a larger number of schools, is necessary before deciding whether this tendency, though small, is significant.

Finally, it does not follow that the correlations of the actual school attainments in English and arithmetic with linguistic background are the same as those shown in Table 3. Although it would seem reasonable to expect intelligence to be independent of linguistic background, it must not be assumed that the actual measures of these used in this investigation are uncorrelated. Spearman's rank correlation between the questionnaire and intelligence test (i.e., the correlation between the first and last rows of Table 2) is, in fact, — .29.\* Thus, in proceeding from the actual to the adjusted school attainments, a component negatively correlated with linguistic background, has been removed.

\* This correlation, in the opinion of the writer, justifies a caution concerning the validity of the intelligence test when used as a comparative measure of ability for monoglot and bilingual children. Of course, as the standard error of the coefficient works out at .24, it could be claimed that the coefficient, as a statistic, is consistent with a parameter of zero (see, however, Lewis, 1959).

Correlations between the actual school attainments and linguistic background may, therefore, be expected to yield negative coefficients of even greater magnitude.\*

#### V.—SUMMARY OF CONCLUSIONS.

(1) Attainment tests of English (comprehension and usage) and arithmetic (mechanical and problem) were given to all the 10-year-old pupils from sixteen primary schools in mixed-language areas in Wales. A non-verbal test of intelligence and a linguistic-background questionnaire were also given. The attainment differences between schools were considerable, though in the case of English comprehension, the over-all dispersion of school means was not significant.

(2) These attainment differences were found to be related to the corresponding differences in intelligence (this relationship being noticeably less in the case of mechanical arithmetic, however). In all the attainment tests the school means tended to increase with intelligence, but not regularly.

(3) The school attainments were adjusted to a common basis, as regards intelligence, by the analysis of covariance. The over-all dispersion of the adjusted school means was significant (at the 1 per cent. level) in all the attainment tests. Differences in intelligence between schools do not, therefore account fully for their attainment differences.

(4) A linguistic background index was formulated to assess the 'Welshness' of background of the school groups. Schools whose attainments in English—adjusted for intelligence—were markedly high, were observed to have low linguistic indices (i.e., low Welsh background), while schools whose adjusted attainments in English were low had high linguistic indices (i.e., high Welsh background). The tendency for high adjusted attainments to be associated with low linguistic indices and for low adjusted attainments with high linguistic indices was confirmed as being strong in both English tests, and was especially strong in English usage. (There is some evidence, too, to suggest that a relatively poor attainment on English usage, as compared with English comprehension, is also associated with a high linguistic index.)

In arithmetic, on the other hand, the tendency for high adjusted attainments to be associated with low indices, and low adjusted attainments with high indices, is only slight.

(5) Linguistic background may thus be regarded as an important factor with respect to those school differences of attainment in English which are not accounted for by intelligence. With regard to the corresponding attainment differences in arithmetic linguistic background is of less importance, further research being necessary to decide whether the slight tendency observed is real.

#### VI.—REFERENCES.

- BARKE, E. M. (1933). A study of the comparative intelligence of children in certain bilingual and monoglot schools in South Wales. *Brit. J. Educ. Psychol.*, 3, 237-250.  
 BARKE, E. M., and PARRY-WILLIAMS, D. E. (1938). A further study of comparative intelligence of children in certain bilingual and monoglot schools in South Wales. *Brit. J. Educ. Psychol.*, 8, 63-67.

\* This, of course, the reader may verify in the case of the English tests by correlating the last row of Table 2 with the ranks of school attainment means read from Figures 1 and II.

- GUILFORD, J. P. (1936). *Psychometric Methods*. New York: McGraw-Hill Co.  
 HOFFMANN, M. N. H. (1934). *The Measurement of Bilingual Background*. New York: Columbia University.  
 JONES, W. R. (1952). The language handicap of Welsh-speaking children. *Brit. J. Educ. Psychol.*, 22, 114-123.  
 LEWIS, D. G. (1959). Bilingualism and non-verbal intelligence; a further study of test results. *Brit. J. Educ. Psychol.*, 29, 17-22.  
 LINDQUIST, E. F. (1940). *Statistical Analysis in Educational Research*. New York: Houghton Mifflin Co.  
 MILES, T. E. (1953). *Bilingualism in Caernarvonshire*. Cardiff: University of Wales Press.  
 NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH IN ENGLAND AND WALES (1947). *A Scale of Non-verbal Mental Ability*.  
 SAER, D. J. (1922). An enquiry into the effect of bilingualism upon the intelligence of young children. *J. Exp. Ped.*, 6, 232-240, 266-274.  
 SAER, D. J. (1923). The effect of bilingualism on intelligence. *Brit. J. Psychol.*, 14, 25-38.  
 SCHONELL, F. J., and SCHONELL, F. E. (1950). *Diagnostic and Attainment Testing*. Edinburgh: Oliver and Boyd.  
 SMITH, F. (1923). Bilingualism and mental development. *Brit. J. Psychol.*, 13, 271-282.

(Manuscript received, January 30th, 1959.)



# A CRITICAL STUDY OF BILINGUALISM AND NON-VERBAL INTELLIGENCE

By W. R. JONES

(*Department of Education, University College of North Wales, Bangor.*)\*

**SUMMARY.** Previous investigations undertaken in Wales have yielded somewhat contradictory results concerning the influence of bilingualism on performance in non-verbal tests of intelligence. More recent surveys, planned on a wider scale, consistently reveal a tendency for mean scores on such non-verbal tests to decrease with increasing Welsh linguistic background, and also show statistically significant differences in favour of monoglot (English) groups by comparison with bilingual (Welsh) groups. In the present study, reference is made to a re-analysis of the results of one of these large-scale surveys in relation to the parental occupations of the pupils tested. The findings of this re-analysis indicate that, whereas monoglot and bilingual groups which vary significantly in occupational class also differ significantly in non-verbal intelligence, corresponding linguistic groups of comparable socio-economic status do not differ significantly in this respect. It is, therefore, concluded that bilingualism need not be a source of intellectual disadvantage. The importance of a thorough examination of socio-economic factors in any comparative study of monoglot and bilingual children is strongly emphasised, and a study of bilingualism and non-verbal intelligence which appeared in a recent issue of this *Journal* is critically examined on account of its defects in this and in certain other respects.

## I.—INTRODUCTION.

THE general trends in the literature dealing with the effects of bilingualism on intelligence warrant the conclusion that, whereas bilingual children suffer from a language handicap in verbal tests of intelligence, there is no evidence for a similar inferiority when their performance on non-language tests of intelligence is measured against that of monoglot children (Rumyanek 1931, Arsenian 1937, and Darcy 1953). It is not surprising, therefore, to find that from about 1930 onwards, researchers in Wales have used non-verbal as well as verbal tests of intelligence in their experimental studies of bilingualism. Barke (1933) found that bilingual children were, on the whole, slightly superior to monoglot children in a non-verbal test of intelligence, whereas in a verbal test administered entirely in English, the bilinguals were consistently inferior to the monoglots. Both monoglot and bilingual groups were thought to compare very fairly from the point of view of parental occupation, although further inquiry suggested that the home environment of the bilingual group was somewhat superior to that of the monoglots. In a subsequent investigation, Barke and Parry-Williams (1938) found similar results, and in addition, discovered that the superiority of the monoglot group was even more marked in a verbal test of intelligence given to each group in its mother-tongue. The conclusions of both investigations, however, are considerably weakened by the complete absence of proper criteria for evaluating the statistical significance of differences in mean scores between the various groups compared. Jones (1933) compared monoglot and bilingual groups in urban areas on a progressive scale of verbality, ranging from non-language and concrete-verbal material to highly abstract verbal material at the upper end of the scale. He found no statistically significant differences between

\* My thanks are due to Professor D. W. T. Jenkins and Dr. J. Rogers who kindly read the typescript and offered valuable and helpful suggestions.

monoglot and bilingual groups on any of the tests within the scale. James (1947) as part of an enquiry into the general educational performance of monoglot and bilingual groups administered Form A of a non-verbal test of intelligence in English to both groups, followed by Form B administered in Welsh to the bilingual group and in English to the monoglot group. Group differences in mean I.Q.s. were then evaluated by the *t*-test of significance, and the results showed that there was no significant difference between monoglot and bilingual groups of similar social background on either administration of this non-verbal test. Jones and Stewart (1951) were mainly interested in the influence of the verbal factor on the comparative performance of monoglot and bilingual groups in intelligence tests. They found highly significant differences in favour of the monoglot group on both non-verbal and verbal tests of intelligence. The difference in favour of the monoglot group in 'verbal intelligence' remained at a highly significant level even after adjustment had been made by the methods of analysis of covariance to allow for the initial difference between the groups on the non-verbal test of intelligence. It was, therefore, concluded that the difference in adjusted verbal scores was probably due to the verbal factor in intelligence tests, which had seemed to work against the bilinguals in relation to the monoglots. The investigations of Jones (1933), James (1947) and Jones and Stewart (1951) show considerable advances, not only in the application of statistical techniques to the bilingual problem, but also in the adoption of various methods for quantitatively assessing the linguistic background of the children tested.

## II.—RECENT SURVEYS WITH NON-VERBAL TESTS OF INTELLIGENCE.

The somewhat contradictory findings of earlier investigations with non-verbal tests have stimulated subsequent researchers to undertake surveys of the problem on a wider scale than was previously possible. Three recent surveys may be mentioned :

(1) In 1951 the Collegiate Faculty of Education of the University College of North Wales, Bangor, undertook a survey in Caernarvonshire schools, when all children between the ages of 10 years and 11 years 11 months were given a series of tests. The main purpose of the survey was to standardise two intelligence tests on groups of children of varying linguistic background. One of these tests was non-verbal in content, but administered with oral instructions which had been adapted into Welsh for use with Welsh-speaking children. Particularly important for the work of standardisation was the Language Questionnaire, which had been specifically prepared in connection with the survey along the lines of a well-known Bilingual Schedule (Hoffman, 1934). Details concerning its construction, administration and methods of scoring have been published elsewhere (Miles, 1954). By combining Language Questionnaire Ratings and Head-teachers' Estimates of children's linguistic background, pupils were eventually classified into four linguistic groups as follows : (a) Welsh, (b) Mixed-Welsh, (c) Mixed-English, and (d) English. Separate statistical analyses were then carried out for the 10-year and 11-year levels. A striking feature in the results was the tendency for mean raw scores on the non-verbal test to increase as the linguistic composition of the group became progressively English. The over-all differences between the group means were also highly significant at both age levels.\* More detailed comparison of the various group means by an application of the *t*-test of significance revealed highly significant differences in

\* The respective values of *F* were 11.89 (*df* = 3 and 1,215),  $P < 0.001$  and 21.36 (*df* = 3 and 1,294),  $P < 0.001$ . (Appendix Tables C and D, Jones 1955).

favour of the English and Mixed-English groups by comparison with the Welsh group, and also a significant difference in favour of the English group by comparison with the Mixed-Welsh group. These differences were equally present at both age levels. In addition, at the 11-year level a highly significant difference was found in favour of the Mixed-Welsh group by comparison with the Welsh group.\* A full account of this survey may be found in publications sponsored by the Collegiate Faculty of Education at Bangor (Jones, 1955 and 1959).

(2) In 1954 a panel of four researchers representative of the Collegiate Faculties of Education at Aberystwyth and Bangor investigated the educational attainment of monoglot and bilingual children between 10 and 11 years of age in relation to their intelligence and linguistic background. This survey was undertaken at the request of the Welsh Joint Education Committee which later published a detailed account of the investigation (Jones *et al.*, 1957). In all, 749 children were tested in primary schools in the counties of Anglesey, Caernarvon, Merioneth, Cardigan, Pembroke and Carmarthen. The non-verbal test of intelligence and the Language Questionnaire which were used in the 1951 Bangor Survey were also administered on this occasion. The children were classified into the following linguistic groups, according to their scores on the Language Questionnaire: (i) Welsh (+50 per cent. to +100 per cent.); (ii) Mixed-Welsh (0 per cent. to +49 per cent.); (iii) Mixed-English (-1 per cent. to -70 per cent.); and (iv) English (-71 per cent. to -100 per cent.). Highly significant overall differences were found between the groups on the non-verbal test.† A recent statistical analysis of the data has also revealed a tendency for mean raw scores on the non-verbal test to increase as the linguistic composition of the group becomes increasingly more English, and further comparison of group means has indicated highly significant differences in favour of English and Mixed-English groups by comparison with the Welsh group.‡ In effect, the results of this survey seem to corroborate the general findings of the 1951 Bangor Survey.

(3) Morgan (1955) made a comparative study of the effects of varying degrees of bilingualism on the performance of children aged 10 to 12 years in certain non-verbal tests of intelligence. A synopsis of this work appeared at a later date (Morgan, 1957). He found a "tendency for scores in the three tests to decrease with increasing Welsh linguistic background" (p. 10). After comparing the mean scores of three broad linguistic groups defined as Welsh, Mixed and English, and testing the mean differences for statistical significance, he concluded that "children with a predominantly Welsh linguistic background, that is, the monoglot and near monoglot Welsh children, are handicapped by their bilingualism in non-verbal tests of intelligence." (p. 14). His investigation,

\* Groups aged 10 : 0 to 10 : 11 :

For English and Welsh groups,  $t=5.304$ ,  $df=1,014$ ,  $P<0.001$ .

For Mixed-English and Welsh groups,  $t=2.773$ ,  $df=783$ ,  $P<0.010$ .

For English and Mixed-Welsh groups,  $t=2.397$ ,  $df=432$ ,  $P<0.020$ .

Groups aged 11 : 0 to 11 : 11 :

For English and Welsh groups,  $t=4.994$ ,  $df=1,084$ ,  $P<0.001$ .

For Mixed-English and Welsh groups,  $t=4.158$ ,  $df=848$ ,  $P<0.001$ .

For English and Mixed-Welsh groups,  $t=2.234$ ,  $df=446$ ,  $P<0.020$ .

For Mixed-Welsh and Welsh groups,  $t=3.603$ ,  $df=854$ ,  $P<0.001$ .

(Tables 2 and 4, Jones, 1959.)

† The value of  $F$  was 5.46,  $df=3$  and 745,  $P<0.001$  (Appendix C, Table C7, Jones *et al.*, 1957).

‡ For English and Welsh groups,  $t=3.630$ ,  $df=570$ ,  $P<0.001$ .

For Mixed-English and Welsh groups,  $t=2.667$ ,  $df=437$ ,  $P<0.010$ .  
(Table 6, Jones, 1959).



although specifically concerned with the relationship between linguistic background and test-performance, also afforded some evidence for the influence of other variables, such as urban-rural differences and socio-economic or cultural-educational factors.

### III.—THE IMPORTANCE OF SOCIO-ECONOMIC FACTORS IN 'BILINGUAL' RESEARCH.

From the unanimous findings of the three recent surveys which have been briefly summarised in this paper, one is tempted to conclude that bilingualism has an adverse effect on performance in non-verbal tests of intelligence. Such a conclusion, however, should not be finally accepted without further examination of the evidence, particularly with reference to the influence of variables other than the bilingual factor on test performance. For this reason, the present writer has analysed the results of the 1951 Bangor Survey in the light of socio-economic data which had been made available to him through the co-operation of the School Welfare Officers in Caernarvonshire (Jones, 1959). Detailed information was thus obtained concerning the parental occupation of each child tested in the survey, and these occupations were subsequently coded and classified in a form comparable to the one used by the Population Investigation Committee.\* On re-analysing the results in relation to the occupational data, it was found that various linguistic groups which did not vary significantly in occupational class did not differ significantly in non-verbal intelligence.† Such groups were mainly located in the urban areas of the county. On the other hand, highly significant differences in non-verbal intelligence were found between corresponding linguistic groups situated mainly in the rural areas of the county, and such groups were also shown to differ significantly from each other in respect of occupational class.‡ It also appeared that these highly significant differences in non-verbal intelligence between linguistic groups in the rural areas could be adequately accounted for by occupational rather than linguistic variations between the groups, since variations of the latter kind were equally operative amongst the urban linguistic groups which showed no appreciable differences in non-verbal intelligence. It was, therefore, concluded that bilingualism need not be a source of intellectual disadvantage. Finally, this re-analysis of the 1951 Bangor Survey emphasised the importance of a thorough examination of socio-economic factors in any comparative study of monoglot and bilingual children.

A recent study of bilingualism and non-verbal intelligence which appeared in this *Journal* (Lewis, 1959) is signally defective in this and in certain other respects. The study contains unacknowledged extracts from the data of the 1954 W.J.E.C. Survey, already referred to in this paper.§ It is, indeed, regrettable that the article itself contains no reference to the survey as the source of Lewis's data, so that the general impression left upon the reader is that the author is throughout describing and discussing his own original data. The complete data had been previously published under the names of the four

\* *The Maternity Inquiry in Great Britain*. Population Investigation Committee and Royal College of Obstetricians and Gynaecologists. Oxford University Press, 1948.

† Significance of differences in non-verbal test score :

$F = 1.53$ ,  $df = 3$  and  $663$ ,  $P < 0.200$ .

Significance of differences in occupational class :

$\chi^2 = 15.175$ ,  $df = 15$ ,  $P = 0.70-0.50$ .

‡ Significance of differences in non-verbal test score :

$F = 24.78$ ,  $df = 3$  and  $1,719$ ,  $P < 0.001$ .

Significance of differences in occupational class :

$\chi^2 = 132.093$ ,  $df = 18$ ,  $P < 0.001$ .

§ *Supra*, p. 5.

researchers who were solely responsible for planning and directing the whole survey, and Lewis's contribution on the statistical side had been duly acknowledged at the end of this publication (Jones *et al.*, 1957).

Adopting some of the data of the W.J.E.C. Survey, but using a different linguistic grouping, Lewis finds a real difference between monoglot and bilingual children in performance on a non-verbal test of intelligence, and suggests that this difference may be due, in part, to (i) the test being a timed one, and (ii) the groups not being equal with respect to urban-rural differences. He also suggests that the presence of a small but appreciable verbal factor in the non-verbal test as shown by Emmett (1949) could well prove to be far from negligible in assessing the intelligence of the bilingual child.

To suggest certain characteristics of the non-verbal test of intelligence as partial explanations of the statistically significant differences in favour of monoglot children is not altogether convincing, since similar tests have been used in other investigations without any disadvantage to bilingual children by comparison with monoglot children. Indeed, it is particularly relevant to note in this connection that the non-verbal test of intelligence\* as well as the corresponding Welsh version† referred to in the author's study were also used in the 1951 Bangor Survey with definite indications that groups of monoglot and bilingual children of comparable socio-economic status do not differ significantly in average performance on the test (Jones, 1959). There remains the third explanation suggested by the author, namely, the influence of urban-rural differences on test performance, due to "a slightly higher proportion of the Welsh speaking bilinguals coming from a rural environment than of English-speaking monoglots" (Lewis, 1959, p. 21). Studies which provide information about urban-rural differences do, in fact, indicate that rural children are found to be lower scorers than town children, but, when the analysis is carried further and pupils are grouped according to the occupational status of their family, the crude urban-rural differences tend to disappear. According to a recent survey, "a comparison of the urban and rural professional groups reveals that the difference between their mean scores is negligible; a similar comparison of rural and urban manual groups likewise shows little significant difference. However, when the mean score of the combined urban and rural professional groups is compared with that of all those employed in manual occupations, there appears a significant difference with the former achieving the higher mean scores" (Barr, 1959, p. 58). It appears that the wide differences in occupational class which usually exist between urban and rural parts of the country, may be regarded as particularly prominent amongst the various possible causes of urban-rural differences. One would, therefore, expect to find occupational as well as broad urban-rural differences revealed in Lewis's analysis, but, rather surprisingly, it is stated that "the groups of this investigation did not differ appreciably with regard to parental occupation" (Lewis, 1959, p. 21). No firm evidence is offered in support of this statement; on the contrary, the author himself concedes that "a more thorough analysis of the socio-economic factor might have revealed some differences" (Lewis, 1959, p. 21). Such an inadequate treatment of the socio-economic factor is not only unsatisfactory but also difficult to understand in view of the fact that precise information concerning parental occupations was available to the author in the case of each pupil tested in the original survey.

In conclusion, Lewis's discussion of the Language Questionnaire also calls for comment. The possible range of scores on the questionnaire extended from

\* J. W. JENKINS: *A Scale of Non-Verbal Mental Ability* (N.F.E.R.).

† W. R. Jones: *Proffon Dealltariaeth Di-iaith, Cyfaddasiad Cymraeg* (N.F.E.R.).

+100 per cent. (predominantly Welsh) to -100 per cent. (predominantly English), and it was originally decided to separate off a group of children who were definitely Welsh (with scores from +50 per cent. to +100 per cent.), a second group of those who were definitely English (with scores from -71 per cent. to -100 per cent.), and a third group labelled 'Mixed' group (with scores from +49 per cent. to -70 per cent.). Lewis maintains that a Welsh group so defined will show considerable differences in linguistic background, and he consequently suggests a group more narrowly defined with scores ranging from +90 per cent. to +100 per cent. Correspondingly, his English group is restricted to pupils with a range of scores from -90 per cent. to -100 per cent. The remaining children are divided into two groups with scores ranging respectively from 0 per cent. to +89 per cent., and from 0 per cent. to -89 per cent. He also emphasises that the Welsh and English groups so defined are secured without a serious depletion of numbers in each group. Numerical considerations of this kind are not relevant to methods of linguistic classification, which are usually decided upon in accordance with principles related to the general linguistic patterns prevalent in the community. It is also essential that a decision concerning the methods of linguistic classification should be made well in advance of the investigation proper and without any preconceived ideas as to the adequacy of numbers in any given linguistic category. Strong reasons have been offered in support of the original method of classification adopted by the designers\* of the Language Questionnaire and an examination of pupils' answers has indicated the sort of basis on which their decision was made (Miles, 1954). In view of these reasons, other possible grouping plans, including the one used in Lewis's study, were rejected by the four researchers of the W.J.E.C. Survey in favour of the classification originally proposed in connection with the Language Questionnaire. Finally, it should be borne in mind that no single set of questions can adequately reflect the intricate and often contradictory characteristics of a pupil's linguistic background. At best, a Language Questionnaire is but an imperfect instrument, hardly suitable for discriminating between the finer shades of linguistic background. It would, therefore, appear that broad linguistic classifications are often more realistic and consequently more informative for practical purposes than any attempts to separate pupils into linguistically homogeneous sub-groups, artificially defined within excessively narrow limits of questionnaire scores.

\* J. Rogers, T. R. Miles, and W. R. Jones.

#### IV.—REFERENCES.

- ARSENIAN, S. (1937). *Bilingualism and Mental Development*. New York: Teacher's College, Columbia University.
- BARKE, E. M. (1933). A study of the comparative intelligence of children in certain bilingual and monoglot schools in South Wales. *Brit. J. Educ. Psychol.*, 3, 237-250.
- BARKE, E. M., and PARRY-WILLIAMS, D. E. (1938). A further study of comparative intelligence of children in certain bilingual and monoglot schools in South Wales. *Brit. J. Educ. Psychol.*, 8, 63-77.
- BARR, F. (1959). Urban and rural differences in ability and attainment. *Educ. Res. (N.F.E.R.)*, 1, 49-60.
- DARCY, N. T. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *J. Genet. Psychol.*, 82, 21-57.
- EMMETT, W. G. (1949). Evidence of a space factor at 11+ and earlier. *Brit. J. Psychol. (Statist. Sec.)*, 2, 3-16.
- HOFFMAN, M. N. H. (1934). *The Measurement of Bilingual Background*. New York: Teacher's College, Columbia University.



- JAMES, C.B.E. (1947). *A Comparative Study of General Performance between Bilingual and Monoglot Children in South Wales*. Unpublished Thesis, Edinburgh Univ.
- JONES, W. R. (1933). *Tests for the Examination of the Effect of Bilingualism on Intelligence*. Unpublished M.A. Thesis, University of Wales.
- JONES, W. R., and STEWART, W. A. C. (1951). Bilingualism and verbal intelligence. *Brit. J. Psychol. (Statist. Sec.)*, 4, 3-8.
- JONES, W. R. (1955). *Bilingualism and Reading Ability in English*. Cardiff: University of Wales Press.
- JONES, W. R., MORRISON, J., ROGERS, J., and SAER, H. (1957). *The Educational Attainment of Bilingual Children in relation to their Intelligence and Linguistic Background*. Cardiff: University of Wales Press.
- JONES, W. R. (1959). *Bilingualism and Intelligence*. Cardiff: University of Wales Press.
- LEWIS, D. G. (1959). Bilingualism and non-verbal intelligence: a further study of test results. *Brit. J. Educ. Psychol.*, 29, 17-22.
- MILES, T. R. (1954). *Bilingualism in Caernarvonshire*. Cardiff: William Lewis.
- MORGAN, E. R. (1955). *A Comparative Study of the Effect of Varying Degrees of Bilingualism on the Performance of Children aged 10 to 12 years in Certain Tests of Intelligence*. Unpublished M.A. Thesis, University of Wales.
- MORGAN, E. R. (1957). *Bilingualism and Non-verbal Intelligence: A Study of Test Results*. Pamphlet No. 4, Collegiate Faculty of Education, Aberystwyth.
- RUMYANECK, J. (1931). The comparative psychology of Jews and Non-Jews: A survey of the literature. *Brit. J. Psychol.*, 21, 404-426.

(Manuscript received, May 11th, 1959.)

## NOTES AND COMMENTS

### REPLY TO: A CRITICAL STUDY OF BILINGUALISM AND NON-VERBAL INTELLIGENCE (By W. R. JONES)

By D. G. LEWIS.

One important difference between Jones and the writer concerns the selection of groups. Jones approaches this with pre-conceived ideas as to where the group boundaries should be; the decision, he maintains, should be made well in advance of the investigation proper. The writer, on the other hand, maintains that the question should be finally settled only when all evidence to be gained from the investigation is available. In particular, a detailed analysis of the questionnaire data should first be made in conjunction with the teachers' ratings of linguistic background. In an analysis of the questionnaire data from schools in the Aberystwyth Faculty Area, the writer found that the broadly-defined "Welsh" group of the W.J.E.C. Survey contained a substantial minority of children (about 35 per cent.) from homes where Welsh was *not* always spoken. Sometimes only one parent (usually the mother) would be constantly speaking to the child in Welsh, and often there would be brothers and sisters with whom the child would converse mostly in English. Clearly differences of background within a group so defined are very considerable: the group boundaries are excessively wide. And this is particularly unfortunate since it is in a comparison of the two "extreme" groups (Welsh and English) that any effect of a bilingual background can best be assessed\*. Marked heterogeneity within the "extreme" groups is avoided, however, in the grouping adopted by the writer. This grouping, moreover, agreed well with the teachers' ratings. Other tentative groupings, including the one favoured by Jones, did not show the same measure of agreement.

Jones appears to misunderstand the reference to depletion of numbers. These numerical considerations, while not affecting the position of group boundaries, are by no means irrelevant. If by restricting the Welsh group to those of a thoroughly Welsh background, numbers are seriously depleted—so that a group mean score would involve an unduly large standard error—the remedy is not to widen the group boundaries but to test more children. As it happened, however, this was not necessary.

Jones' other two criticisms merit but brief comments. A *thorough* investigation of socio-economic status would involve other variables besides parental occupation. The group comparison of parental occupation merely ensured there were no gross disparities. (Incidentally, the W.J.E.C. Report completely ignores the socio-economic factor.) Finally, the original questionnaire and intelligence-test data had been made available by the Department of Education at Aberystwyth. The writer is glad to acknowledge the generous co-operation of all in the Department interested in this field of research.

#### Comments on Dr. Bene's Article.

By W. B. DOCKRELL

(University of Alberta)

In her article "Suppression of Heterosexual Interest and of Aggression by Middle Class and Working Class Grammar School Boys," in the *Journal* of November, 1958, Dr. Bene concludes that "The present results indicate only that there are no social class differences regarding suppression of aggression." The data included in the article are limited and it is difficult to re-work the material, however, it does not

\* As pointed out by the writer (this *Journal*, 29, 20) this comparison is essentially one between a monoglot (English) group and a bilingual (English-Welsh) group.

seem that this conclusion is supported by the evidence presented. The differences between the two groups are small on some of the items designed to indicate suppression of aggression, but not so small that one would accept the null hypothesis. None of them are large enough for one to reject the null hypothesis, but not rejecting and accepting are two very different matters. This may merely be a slip in wording, but there are no doubt readers who took this clause literally.

Without further information, it is difficult to decide whether the data presented might not be interpreted to support the original hypothesis. One would wish to know rather more about the sample than can be learned from this article. For instance, Dr. Bene refers to four grammar schools but describes only three. A grammar school drawing predominantly working-class boys might convey middle class standard rather less effectively than a mixed or predominantly middle class one. Similarly, one would wish to know more about the sample of modern school boys referred to in the discussion.

The form of the testing instrument itself also makes re-interpretation difficult. What was the precise form of the items used to indicate suppression of aggression? At least two were open ended stories where the endings were classified as non-aggressive in terms of overt behaviour; non-aggressive in terms of thoughts and feelings; and what else? How many of the other endings were aggressive, not completed, non-classifiable or characterised in some other way? The other six items seem to call for a straight agreement or disagreement but how many 'don't knows' were there and were the 'don't knows' evenly distributed between the groups? If there were three or more classifications of responses, aggressive, non-aggressive, don't know and some others, were the differences in numbers of aggressive responses not significant?

In analysing the data, Dr. Bene tested each item for significance individually. From the data given, it would seem reasonable to divide the questions into projective items of feeling and descriptions of preferred behaviours. The two 'projective' items are counted twice because in both cases two different classifications of response are labelled non-aggressive. Can the two groups of non-aggressive responses to the story about Fred be combined and contrasted with the aggressive responses? If we combine these two groups and contrast them with the responses not described in the article, the  $X^2$  exceeds the .05 level for a one-tailed test. The other projective item, the story about Bob, is the only item in which the proportion of non-aggressive responses from the working class boys exceeds that from the middle class boys, so that if the responses from the two projective questions are combined, the difference between the two groups is not significant. On most tests we expect the total score and not the individual items to discriminate. It is not clear from the article whether it would be legitimate to combine items in this case. If, however, we do add the 'behavioural' items, we find a  $X^2$  greater than .025 for a one-tailed test.

In her discussion, Dr. Bene refers to other uses of the same instrument as evidence of its validity. Were any investigations made into its reliability? Were the two groups matched for family size, for this is cited as a relevant variable?

At most Dr. Bene's findings are inconclusive. She failed to find differences as large as, or of the same kind as, she had found between middle-class boys who had several siblings and those whose life evolved mainly round adults—though nine of the ten differences between the social classes were in the direction hypothesised. More extensive testing with the same instrument might yield reliable differences between the social classes, so might a different form of investigation. In the meantime, we must suspend judgment.



## Reply to the Comments by Professor Dockrell.

By EVA BENE

I would like to start my reply to Professor Dockerell's comments by stating that I agree that an investigation into social class differences in England which uses different techniques from those I used, might yield different results. We can never be certain that research results correspond with reality but only believe with a greater or lesser degree of confidence that they do. I hope that not many of my readers have understood me to claim that the fact that I did not find a significant difference between middle class and working class grammar school boys regarding suppression of aggression definitely proves that such a difference does not exist between the social classes in England.

The objections Professor Dockrell raised against my paper are to some extent semantic in origin. They partly hinge on the meaning of the word 'indicate.' While I used this term in the sense of 'suggest,' he appears to have understood it in the sense of 'prove.' It would be interesting to know which meaning is more frequently attached to this word when it is used in the phrase: "The results indicate." Another difficulty of a semantic nature has arisen from the circumstance that Professor Dockrell attached great significance to the meaning of half a sentence taken out of its context. His main objections are to the following phrase: "The present results indicate only that there are no social class differences regarding suppression of aggression." This sentence continues in the following manner: "... they do not show to what extent both groups of boys express or suppress their aggressive feelings." Professor Dockrell thinks that the phrase which he quotes is misleading. Maybe it is, but so is taking it out of its context which gives its meaning a different slant, and so is ignoring the next sentence which begins with the words: "It seems . . ."

Now to the objections on technical grounds. Professor Dockrell is correct in saying that I referred to four, but only described three grammar schools. The sentence describing the grammar schools should have read: "One school was located in a predominantly working class district, one in a predominantly middle class district, and two in districts in which both social classes were about equally represented." Unfortunately, part of this sentence was left out, but fortunately, the paper contained references which should enable any interested reader to find out more about the sample.

Professor Dockrell's next objections have to do with the story completion items. In the case of these items each subject had to give two separate responses. He first had to say what he thought the hero of the story did in the situation, and then what he imagined the thoughts and feelings of the hero were. For some reason, Professor Dockrell thought that the responses to each of these items were counted twice when it was not legitimate to do so. Yet he combined them and worked out a Chi square on the basis of the combined responses. Since he only had the percentage figures published in the article at his disposal, it is quite mysterious how he went about his calculations, and the Chi square he gives appears to be meaningless. The same holds for his calculations of the other Chi square. These Chi squares are meaningless not only for statistical reasons but also because he obtained them by combining items which are different in kind. He could have seen from the table that some of the 'behavioural' items were statements and others were not.

Professor Dockrell also states that "... nine of the ten differences between the social classes were in the direction hypothesized," whereas it can be seen from the table that one of the ten comparisons showed no difference, seven went in the predicted and two went in the other direction.

However, Professor Dockrell also raises some interesting points. He would like to know how many of the non-aggressive responses were aggressive, how many non-classifiable, how many not completed, etc. I can understand that a reader who is really interested in a piece of research would like to see all the data. The reason why it is not given is that this would result in papers which are too long and contain

too many tables for publication. He also asks whether any investigations were made into the validity of the items used. The question of the validity of the instruments used is too complex to be answered in a sentence or two. All I can give as an answer is that it is almost impossible to validate psychological tools which have to do with the investigation of emotions, at least to validate them in a straightforward manner. The reason for this is that there are no valid external criteria available with which one could compare the obtained results.

Some psychologists believe that for this reason it is wrong to try to investigate emotions. Others believe that we should proceed to collect information about the emotional lives of people the best we can.

To another question Professor Dockrell asks, there is a simple answer. He would like to know whether the two groups investigated were matched for family size. Yes, they were.

Finally Professor Dockrell raises the question whether it is proper in research to expect individual items instead of total scores to discriminate. The answer again is yes. Both methods have their advantages and disadvantages. Some of the advantages of working with separate items are the following: One does not lose sight of the results obtained by individual items. One can use different types of items and thereby eliminate the danger that the responses are too much influenced by the manner in which the items were phrased. One eliminates the difficult and often unanswerable question of how much weight should be attached to the different responses. Regarding the interpretation of results: one can draw conclusions on the basis of the pattern and the significance of the differences between the responses of two groups to a number of items just as one can on the basis of the difference between the means of two groups.

I would like to finish my reply to Professor Dockrell by stating once more that I believe as firmly as he does that the fact that in an investigation no significant differences were found does not prove the null hypothesis to be true.

## RESEARCH NOTE

### The Peabody Picture Vocabulary Test with English Children.<sup>1</sup>

By JAMES W. MOSS, Ph.D.<sup>2</sup>

(*Medical Research Council, Maudsley Hospital, London*)

AND PHYLLIS EDMONDS

(*Institute of Education, University of London*)

The present study was conducted to provide preliminary data on the utility of the Peabody Picture Vocabulary Test<sup>3</sup> (PPVT) with English children. The PPVT is an instrument recently developed in the United States which purports to measure comprehension of the spoken word, including the ability to associate verbal symbols with pictorial representations of objects, events, and actions. The test consists of 150 plates arranged in ascending order of difficulty. Each plate contains four pictures: three decoys and a correct one. The same set of plates is used for both forms of the test. The test is administered by having the child indicate which picture on the plate best illustrates the meaning of the orally presented stimulus word. No particular form of response is required. The subject is tested from a basal of eight consecutive correct responses to a ceiling of six errors out of eight consecutive responses.

The PPVT was standardized on random samples of normal children from 2 years 6 months through 18 years of age, but it was specifically designed to be useful with handicapped subjects. The test is not timed. The comparatively large drawings make it of value with some partially-sighted children. It may also be used with cerebral palsied and speech-handicapped children.

The PPVT was administered to 101 English school children between the ages of six and nine from two different London County Council schools. The Otis Group Intelligence Scale, Primary Examination, was administered at the same time. It was found, with the age range covered, that the PPVT basal and ceiling levels of the children generally fell between plates 41 and 90, inclusive. For the item analysis, therefore, only the responses to these plates are considered. Only those children in the sample with I.Qs. 90 to 110, inclusive, were involved in the cross-cultural comparison of PPVT scores. The sample was divided into two age groups in order to increase the reliability of the results. The high age group ( $N=21$ ) ranged in CA from 7:6 to 8:5 with a mean CA of 8:0. The low age group ( $N=17$ ) ranged from 6:3 to 7:5 with a mean of 7:2. No differences were found between the mean scores or the variances of the English and American children at either age level. The correlation between the PPVT and the Otis Scale was 0.68. With the effects of CA partialled out, the correlation was reduced to 0.64.

The pictures in the PPVT are, in general, appropriate for English children, although there are a few drawings, such as the American policeman, baseball bat, or coins which might cause difficulty. Some of the words, or word-picture combinations are inappropriate resulting in a change in the difficulty level for the particular item. The present arrangement of the items is not wholly appropriate for use with English children as it does not represent a smooth, ascending order.

The general conclusions reached by the present authors is that the PPVT is potentially a useful instrument for use with English children. Although it can evidently be used in its present form with the American norms, it should be re-standardized for English usage.

<sup>1</sup> This project was supported in part by Research Grant MF 8599 from the National Institute of Mental Health, Bethesda, Md. An extended report of this study may be obtained without charge from James W. Moss, San Francisco State College, San Francisco, California.

<sup>2</sup> Now at San Francisco State College, San Francisco, California.

<sup>3</sup> Dunn, L. M. Manual, Peabody Picture Vocabulary Test, Tentative Edition, George Peabody College for Teachers, Nashville, Tennessee, 1959.



# SUMMARIES OF RESEARCHES REPORTED IN DEGREE THESES

## Sources of Fluctuation in Scores from Successive Group Intelligence Tests.

By NORMAN W. KING

*(Abstract of a thesis submitted in part-fulfilment of the requirements for the Degree of Bachelor of Education at Aberdeen University, in 1959.)*

### AIM.

To examine the incidence of discrepancies of ten or more points of I.Q. from two parallel group tests administered within an interval of one week, and to discover the proportion of such cases where internal evidence from the test booklets indicates one or other as the 'correct' score.

### SUBJECTS.

In a year-group of 2,768 Aberdeen children aged 11, there were 194 cases where discrepancies of ten or more points were found between scores from two successive intelligence tests (after practice effect had been removed by re-standardizing the scores in each test to a mean of 100 and a standard deviation of 15 in the year-group).

### TESTS.

Moray House Verbal Reasoning Tests 55 and 56 were given, with one week's interval; previous practice and limited coaching having been given on Moray House Test 49.

### METHOD.

The 194 cases were classified from a study of the test booklets into four groups:

- (a) Ceiling cases: where raw score on both tests was over 90 out of 100 items, and the discrepancy arose through inadequacy of the standardization at the highest limits (20 per cent.).
- (b) Slow performers: where one of the two tests was unfinished, at least fourteen items being left unanswered (8 per cent.).
- (c) Block errors: where, amid a run of correct items, a whole section was wrong because the instructions had been misunderstood (17 per cent.).
- (d) The remainder (55 per cent.).

### RESULTS.

The discrepancies of 10 points or more in 108 boys and 88 girls (7 per cent. of the group), constituted a higher incidence than would have been predicted from the reliability coefficients of the tests (slightly over 3 per cent.). There were eight cases of discrepancies of 20 points or more. Nine-tenths of the discrepancies showed an I.Q. (mean of the two scores) above the group mean of 100.

The large proportion of ceiling cases was partly the result of the accumulation of practice and coaching, the mean score in the tests before re-standardization being 104.9 in the first and 108.4 in the second test.

An attempt was made to determine the 'correct' score for group (d), where there was no clear internal evidence of the source of the discrepancy, using as a criterion the mean of scores in English and Arithmetic tests and scaled teachers' estimates. In half the cases, the mean of the two I.Qs. came closest to this criterion; in the other half, in approximately equal numbers, the higher or the lower score was nearer to the criterion.

### CONCLUSION.

Where discrepancies of 10 or more points occur, it is recommended that the test booklets be examined. For ceiling cases, slow performers and block errors, the higher score should be accepted, and this would deal with nearly half the cases. For the remainder, the mean of the two scores must be taken as the best available estimate.

# A Factorial Study of the Relationship between the Child's Vocabulary and His Reading Progress at the Infants' Stage.

By ERIC POTTS

(Abstract of thesis submitted for the degree of M.Ed. in the University of Manchester, April, 1959.)

## AIMS.

In the literature concerned with the preparatory stages of reading, teachers are often urged to take steps to extend their children's vocabularies. The authorities aver that the child with the more extensive vocabulary is better equipped to tackle the printed word. However, there is little published evidence obtained from children in the early stages of reading, to support this view. The aims of this research were, therefore :

- (1) To test the hypothesis that an extensive vocabulary predisposes a child to early success in reading.
- (2) To analyse the factors underlying a battery of tests chosen for their assumed relevance to the measurement of reading attainment and vocabulary knowledge.

## THE SAMPLE.

One-hundred-and-seven boys and ninety girls, aged between 72 and 78 months, drawn from eight infants' schools, performed all the tests of the present battery. The distribution of intelligence quotients in this sample agreed closely with the normal frequency curve, and their parents' occupations were typical of an urban area of South-East Lancashire.

## METHOD.

A multiple-choice pictorial vocabulary test was constructed in order to obtain measurements of the comparative sizes of the children's vocabularies. It was assumed that this estimate of vocabulary extent would be directly proportional to the absolute vocabulary of the subject. A specially adapted sampling technique had to be employed for selecting the items, as only words from certain Thorndike frequency categories were considered suitable for the age group. This test, presented in two parallel forms, was found to have a reliability of .923. Two other tests were constructed, one to measure the child's ability to discriminate visually, and the other, composed mainly of nonsense words, to assess ability to read regular phonic units.

The complete battery of tests and the assessments used were as follows :

- (1) Chronological age.
- (3) The Burt-Vernon Word Reading Test.
- (4) A pictorial test of Visual Discrimination.
- (5) The Moray House Picture Intelligence Test (Mental Age).
- (6) Net school attendance.
- (7) The Schonell Silent Reading Test A.
- (8) A test of 'Phonic' ability.
- (9) Social and Cultural Environment rating.

## STATISTICAL ANALYSIS.

The differences in means in reading score (Burt-Vernon), vocabulary score and mental age, for school and sex groups were tested for significance by means of the 't' ratio.

The inter-correlations of the tests and other assessments were calculated by the product-moment formula. The resulting correlation matrix (Table 1) was submitted to factorial analysis by the centroid method.

TABLE 1  
THE CORRELATION MATRIX.

Variable	2	3	4	5	6	7	8	9
(1) Chronological Age....	.140	.324	.036	.191	.608	.294	.233	.008
(2) Vocabulary Score....		.342	.372	.640	.035	.440	.430	.309
(3) Burt-Vernon Reading Score.....			.375	.573	.216	.757	.787	.450
(4) Test of Visual Discrimination.....				.548	-.077	.443	.330	.279
(5) Moray House Mental Age.....					.056	.590	.556	.453
(6) Net School Attendance.....						.149	.095	-.204
(7) Schonell Silent Reading Test A.....							.783	.431
(8) Test of Phonic Ability.....								.414
(9) Environmental Score.....								—

The extracted factors were tested for significance by four independent criteria, and subsequently rotated until interpretation became possible.

#### THE RESULTS.

Highly significant differences were found among the school means for the Burt-Vernon reading test. These were accompanied in a few cases by parallel differences in school means for mental age, but another important cause emanated from variations in school policy on the introduction of 'phonics.' As the Burt-Vernon test appears to favour children with considerable phonic skill, the schools where phonics were introduced early in the reading programme achieved higher mean scores on this test. Moreover, the fact that the children in these schools were able to 'read' by purely mechanical means reduced their reliance upon understood vocabulary. Thus, the correlations between vocabulary score and Burt-Vernon reading score, in the cases of these schools, ranged only from .3 to zero. In contrast, for the schools in which phonics are deferred to a relatively late stage in the reading programme, the corresponding coefficients were of the order of .6.

The full sample correlation coefficients between vocabulary score and reading scores on the Burt-Vernon and the Schonell tests were .34 and .44, respectively, despite the intrusion of the purely mechanical means of reading.

Three significant factors were obtained from the correlation matrix. After rotation these were identified as follows:

- I 'g,' which contributed over half the common variance of the test battery.
- II 'Maturity-schooling,' which was concerned mainly with the two measurements of time, viz., chronological age and net school attendance.
- III A factor confined to the three tests of reading, given the title 'phonic reading.'

The rotated factor loadings which are greater than three standard errors of a Pearson zero coefficient are given in Table 2.

TABLE 2  
SIGNIFICANT LOADINGS—ROTATED FACTORS.

Variable	I	II	III	$h^2$
(1) Chronological Age.....	—	.787	—	.658
(2) Vocabulary Score.....	.684	—	—	.471
(3) Burt-Vernon Reading Score.....	.556	.230	.666	.805
(4) Test of Visual Discrimination.....	.575	—	—	.343
(5) Moray House Mental Age.....	.927	—	—	.871
(6) Net School Attendance.....	—	.771	—	.596
(7) Schonell Silent Reading Test.....	.639	—	.566	.760
(8) Test of Phonic Ability.....	.567	—	.647	.749
(9) Environmental Score.....	.440	—	.357	.360



The vocabulary test's common variance was contributed only by the 'g' factor. The relationship of this test with the two standard reading tests was found to be due to the common influence of the 'g' factor. 'Phonic reading' also made a significant contribution to both of the reading tests, but, as this factor is not strictly orthogonal to the 'g' axis, the indications are that reading attainment is more dependent upon general mental ability than the loadings of the reading tests on the 'g' axis suggest.

Thus, the evidence indicates that, at the six-year-old level, reading attainment and vocabulary extent are more in the nature of different manifestations of general mental ability than distinct, mutually contributive abilities. However, this finding does not necessarily detract from the value of including in pre-reading programmes activities directed towards increasing the child's vocabulary.

### Factors Influencing Musicality.

By P. N. P. WILLIAMS

(Abstract of a thesis submitted for the degree of B.A. at the University of Reading, July, 1959.)

The aim of this investigation was to determine whether senior 'approved' school boys differed from non-delinquents in their degree of musicality or feeling for music. Music may well afford an important outlet for sublimated anti-social drives. It would seem likely, therefore, that unmusical people, who have this channel for sublimation restricted, would be more prone to delinquency.

It was first necessary to find a measure for musicality. Wing in his "Tests of Musical Ability and Appreciation," lists at least fifty tests which have this aim in view. On close inspection, however, these tests would appear to measure either memory, intelligence, musical knowledge or aural acuity. Apart from one or two tests in certain batteries, for example, Wing's, the emotional aspect of music is ignored. It is hardly surprising that these tests have been vigorously criticised by musicians. Lowery in his book *The Background of Music*, states that Mursell, and others, obtained little evidence of correlation between performance in the Seashore tests and musical talent as revealed in school or college work. Mursell emphasizes that music is a unity and that analysis incidental to psychological testing destroys the music.

It was decided that a new test must be constructed which would overcome these deficiencies. As a piece of music is a 'gestalt' it was decided to use actual works by composers of acknowledged worth. Twenty extracts were chosen from such works and each item was performed twice on a piano and recorded.

One of the two performances was made as musically acceptable as the ability of the author as a performer allowed. The other performance was made deliberately less 'musical.' For example, the performance might have been 'wooden,' accents might be misplaced, or part playing obscured. After hearing each pair of performances, the subjects were asked to choose which they 'liked better.'

The test was administered to ten professional musicians. In seven items two of the ten musicians made an 'incorrect' choice, in five items one made an incorrect choice, and in the remaining eight items, no 'unmusical' performances were chosen.

This group had a mean score of 18.1 and a standard deviation of 2.05. These ten musicians were matched with ten non-musicians of comparable age, sex and education. The non-musicians had a mean score of 13.4 and a standard deviation of 2.29.

Fifty-four subjects were also asked to attempt to identify a disguised version of the National Anthem. The mean test score of forty subjects who succeeded in this task was 16.00 as compared with a mean score of 11.64 for the fourteen subjects who failed to identify the music. These results were subjected to a 't' test and a value for 't' of 5.86 was obtained. This is significant at the 0.1 per cent level.

Subjects taking the test were also given A. W. Heim's A.H.4 intelligence test and asked to fill in a questionnaire dealing with their musical background. They were

asked for example, "How often do you go to classical concerts?" From this questionnaire a musical background rating was obtained for each subject.

As a measure of reliability twenty-two subjects, chosen at random, were retested after two weeks and their scores on the two occasions were found to correlate to the extent of  $r=0.81$ .

#### RESULTS.

Using normal subjects of ages between 14 and 68, the scores on the test ranged from 4 to 20. The mean was 11.73 and the standard deviation 3.43.

Table 1 shows product moment correlations.

TABLE 1  
n=127

	Test Score	M.B.R.
Test Score .....	—	—
Musical Background Rating .....	.507	—
Intelligence .....	.231	.283
Age .....	.324	.513

There was no significant difference in scores between the sexes.

#### THE DELINQUENT AND CONTROL GROUPS.

A group of thirty-nine senior 'approved' school boys were compared with the same number of controls of similar age, sex and intelligence. The scores of these two groups are shown in Table 2.

TABLE 2

	Delinquents	Control Group
Mean Test Scores .....	9.95	10.44
Standard Deviation of Test Score .....	3.113	2.318
Mean M.B.R. ....	6.724	3.820
S.D. of Musical Background Ratings.....	3.441	3.136

A 't' test was carried out on the test score results and a value was found for 't' of .781, which is not significant.

A 't' test on the musical background ratings of the two groups gave a value for 't' of 3.845 which is significant at the 0.1 per cent. level.

#### PREFERENCE FOR THE SECOND PERFORMANCE.

Nearly all subjects showed a preference for the second performance, although an equal number of first (A) and second (B) performances were 'correct.' This preference was not, however, shown by the ten musicians or by the retest group.

TABLE 3

Group	"A" Choices	"B" Choices	Chi Squared
10 Musicians .....	103	97	.09
39 Delinquents .....	306	474	18.1
39 Controls .....	291	489	25.1
88 Normals .....	707	1053	34.0
22 Retests .....	218	222	.018

Chi squared is not significant for the Musicians and Retest group. On the other hand, the preference shown by the Delinquents, Controls and Normals for the second performance is significant at the .1 per cent. level.

#### DISCUSSION OF RESULTS.

Perhaps the main outcome of this investigation has been to show the possibilities of a test of musicality constructed on these principles. A pilot study conducted by the author showed that the differences in the two styles of performance had to be comparatively coarse if valid results were to be obtained.

Table 1 shows that intelligence has little association with musicality as measured by the test. This is contrary to Spearman's findings.

The correlation of  $r = .507$  between test score and musical background rating would appear to give support to the teaching of musical appreciation in schools. It may well be, however, that people with a predilection towards music ensure that their children have a good musical background.

The preference for the second performance may be due to the fact that the subjects are already familiar with the music on hearing it for the second time. This hypothesis is supported by the fact that this preference is not shown by the musicians or retest group who would already be familiar with the music.

It might be expected that the higher musical background ratings of the delinquents would be reflected in a higher score on the test but this does not occur. The delinquents were asked to fill in the questionnaire as though they were still at liberty. Their higher musical background rating was completely unexpected by the investigator. It included all forms of music—classical as well as Jazz. It might be argued that the delinquents were trying to please the investigator by claiming to be more interested in music than was, in fact, the case. But, judging by the questions they asked about filling in the questionnaire, this did not seem to be so. Again the only item on the questionnaire where their total score was less than that of the controls was the question relating to private music lessons. This would appear to reflect a difference of economic status between the two groups.

At present it would appear that more research is required to confirm this interesting result and to find for it an acceptable explanation.



## BOOK REVIEWS

BURT, C. (1959). *A Psychological Study of Typography*. Cambridge University Press, pp. xix+67, 15s.

This book directs attention to the part played by modes of printing and page arrangement in the broad field of ideational communication. The author is first concerned with problems of type legibility both among children and adults. Continuing his work with London school children, Sir Cyril investigates certain characteristics of type, including size, form and boldness of letters, width of line, spaces between words and leading: unfortunately, though, no evidence is available on the effects of type on prolonged reading. However, a useful table of recommendations for school text-books is presented. Later, a study of aesthetic typographical preferences in adults is described, including correlations of these preferences with those shown by the same persons in allied fields of pictorial and literary presentation.

The author draws attention to the provisional nature of experimental results based on small numbers of persons. But psychological statisticians will note, with considerable interest, the complementary use of P and T-techniques on a single set of data, finding here a most useful example of the reciprocity principle. It is a matter of conjecture whether clearer results might not have been obtained by separate factor analyses in two groups representing 'arts' and 'science' interests. For the small general factor in the T-technique is an awkward partner and tends to obscure the main structural pattern.

Educational psychologists will appreciate that here is groundwork for an aspect of research frequently assumed in group comprehension tests, and thought could well be given to the extension of these ideas. Nevertheless, confusion could arise between the relative effects of type, reading ability, general ability and the ability to comprehend (in the sense of answering questions following a passage). In this study groups of children are matched for general intelligence and reading ability, and tested by means of comparable comprehension test material, in order to establish the parts played by the legibility of different kinds of page type and arrangement. How far the ability of each individual to comprehend (in the above sense) has been structured and included in this work is a matter which merits careful thought for future investigators.

D. M. LEE.

JONES, W. R. (1959). *Bilingualism and Intelligence*. University of Wales Press, pp. 67, 4s. 6d.

Mr. Jones's monograph brings out the important point, which had been insufficiently realised by most previous workers on problems of bilingualism, that the apparent inferiority of bilingual to monoglot children or adults in (non-verbal) intelligence is often attributable to the tendency of monoglot families to be of rather lower socio-economic status. He provides a useful and critical survey of the large number of researches that have been carried out in Wales into these problems, and demonstrates from his own data the effects of this socio-economic factor.

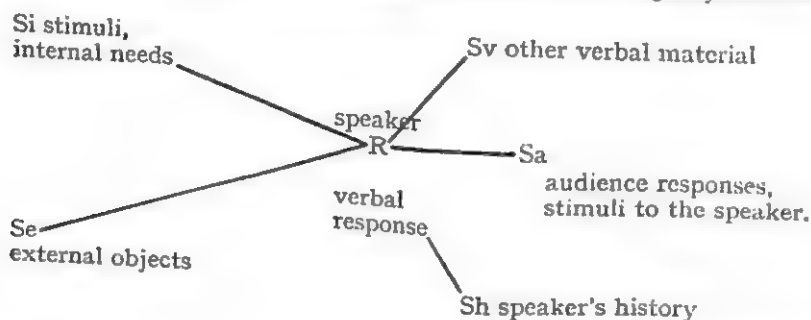
SKINNER, B. F. (1959). *Verbal Behavior*. London: Methuen, pp. x+478, 42s.

B. F. Skinner's latest book is as complete a behaviourist account of the individual's verbal life as one could imagine. In order to achieve this *tour de force* he has stretched traditional behaviourism well beyond some of its earlier limits, as in his stimulating account of 'autoclitic' or self-commenting verbal behaviour required to account for such troublesome elements as intention, attitude, assertion, composition and thinking. Quite apart from its qualities as analysis of verbal behaviour, the book reveals its author's intimate knowledge of a vast range of literary material.

His technique is to describe the behaviourist model in terms solely of a speaker responding by his utterances to internal or external stimuli in the presence of a reacting listener, through whose reactions language is built up in the speaker.

After illustrating the model, he then studies the action of all sorts of variables upon the verbal responses of the speaker drawing upon literature and speech in different settings.

As a thoroughgoing behaviourist he rejects suggestions that we need to invoke the independent entities of 'meaning,' 'idea' or 'image.' A word is a human habit. Verbal behaviour can be explained like other behaviour by the principle of contingency of stimulus, response (by the speaker), reinforcement (by the listener or audience). Speech begins by noises that are reinforced to certain stimuli. These stimuli are internal needs at first and give rise to demanding speech (*Mands*). There are also other verbal responses—*echoic*—*textual* and *intraverbal*. These responses are made to other verbal material. Descriptive responses to objects revealing contact with the outside world are called *tacts*. Finally, *tacts* may become extended by which a speaker responds in a consistent way to common properties of similar objects as exemplified in the *tact*, *pen*. This allows for the richness of vocabulary, for concept formation, metaphor, solecism and the like. The extensions obviously incorporate the speaker's past history. The fundamental condition in verbal behaviour relates to the stimuli which gain control of the speaker's response. Diagrammatically, we may extend Skinner's three-term contingency as follows :



The *mand* accounts for the response under control of the stimuli *Si*, the *tact* when it is under control of *Se*.

Why should we not complete the scheme of control by defining verbal responses controlled by the audience as *auds*, those by the speaker's previous history as *extends* and those by other verbal materials as *reverbs*?

So far, Skinner has assumed, in good behaviourist tradition, that the speaker is merely a response. This is not good enough for most of human verbal activity and in the fourth part of the book he takes up the higher forms of verbal activity shown in the speaker's manipulation of his verbal responses. In the fifth section he looks at editing mechanisms and the productive process of thinking.

The speaker knows what he is saying. Part of his behaviour controls other parts. There is a difference between the response (*tact*) *chair* and the assertion *This is a chair*. The verbal elements that express these differences he calls *autoclitics*. Composition is the binding together of *tacts*, *mands* and other simple behavioural responses by autoclitic devices.

Somewhat similarly in editing and thinking, the speaker reacts as a listener to his own responses.

Hence, to account for this higher activity, Skinner assumes firstly a listener who possesses "at least two systems of responses, one based upon the other," giving rise to autoclitic activity, and secondly, a speaker who can react to himself as a listener, giving rise to editing and thinking processes.

We must ask what appeal this will make to educational psychologists and teachers. What kind of a reader will this material attract? First, in a general background sense it forms a fresh and vigorous re-opening of many problems met in the school learning of language. Its fully behaviourist interpretation, particularly in the first three parts of the book, so well epitomised in the two quotations on pages

44 and 45, will challenge and stimulate readers even if they are not prepared to accept this wholesale throwing over of such linguistic currency as 'ideas' and 'meaning.' In asserting a standpoint that depends almost entirely upon learning, the book brings home to the reader the need always to be on the look-out for the inculcation of correct verbal habits.

As a more specialist text, it has much of use in it for the students of linguistics, semantics, logic and English language. These readers, particularly if they know something about behaviourism, will probably have to ask themselves in the last two parts of the book how far Skinner has been able to adhere to the behaviourist limitations which mark the work of the experimentalists like Hull, Spence and Skinner himself in earlier writing.

The specific and more technical help that this book might give to the educational psychologist and practising teacher is, perhaps, more limited, for many of the qualities of composition, editing, speech and other forms of verbal activity appear in more traditional writing on this subject; but for all that, Skinner's breadth and genius in culling so much material from literature reduces the force of this last criticism. At first the book may strike the reader as being so fresh in its approach as to be difficult. If, however, he takes good care to grasp the essential behaviourist model as set out in the first few chapters of the book and as outlined in the earlier part of this review, he will find that much of the apparently more elaborate material falls easily into place. The work as a whole is an original and stimulating contribution to the study of language.

E. A. PEEL.

THORNDIKE, R. L., and HAGEN, E. (1959). *Ten-thousand Careers*. New York: John Wiley.; London: Chapman and Hall; pp. x+346, 68s.

Vocational psychologists seldom have the opportunity to compare the mental tests they use with the subsequent success of a sufficiently large group employed in any one occupation. However, Professor Thorndike has followed up some 10,000 Air Force recruits who had taken a battery of twenty varied tests twelve years previously, and has obtained information on their present occupations together with certain indices of success. The book provides information on over a hundred occupations which had been entered by about forty or more men each, and analyses the relevance to each of these not only of test scores but also of answers to a biographical data blank.

It emerges that the tests and biographical items gives no reliable predictions of any of the (admittedly unreliable) criteria of success. This is disappointing, but it is in line with much other evidence. There were, however, distinctive patterns of average scores and biographical data among those engaged in different occupations, which make quite a lot of sense. Further analysis, by multiple discriminant function, was carried out on twenty-two jobs, chosen so as not to differ greatly in occupational level. The main 'job-dimension' thus found obviously refers to mathematical-scientific *vs.* non-mathematical tendencies (e.g., physician and mechanical engineer *vs.* hotel manager). The subsidiary dimensions are more obscure though they seem (with a little rotation) to resemble the types distinguished by the reviewer elsewhere: Gregarious *vs.* Isolated and Verbal *vs.* Active.

The vocational counsellor will, of course, wish to study the patterns of scores and background information in detail, and the book is well designed to help him. He will not get a great deal of definitive guidance from it, but at least it gives him some objective indications where previously there was nothing but unverified suppositions.

P. E. VERNON.

TORGERSON, WARREN S. (1958). *Theory and Methods of Scaling*. New York: Wiley and Sons; London: Chapman and Hall; pp. 460, 76s.

Prepared at the request of the Committee on Scaling Theory and Methods of the Social Science Research Council, in the U.S.A., this book presents a very comprehensive and competent synthesis of material previously only available in isolated articles or special chapters of books devoted to the subject matter of a particular



field. As the author phrases it: "In all too many cases there has been too little communication between and within the various disciplines in developing, naming and using the various techniques (of scaling). In some cases the same techniques are used in different areas under different names." The book grapples with the difficult problem of classifying such methods and goes far in the development of a frame of reference, as well as giving an extensive bibliography of about six hundred items of the type referred to in the first sentence. The price alone makes it unsuitable for recommendation as a text-book, but as a work of reference or item for the library of the specialist in problems of measurement it is to be highly recommended as a scholarly and thorough piece of work.

The book begins with a philosophical discussion of the rationale of measurement and scaling, and in the introductory chapters the author develops definitions of three classifications within which the methods of scaling of psychological attributes may be deduced to fall, namely, the "subject centred approach," the "stimulus centred approach," and the "response approach." He makes the point that they fall in a common category, not necessarily outside the field of most scaling in the field of natural science insofar as he rejects scales which depend upon "measurement by definition or fiat," in which an *assumption* of relationship between certain observations and the concept under consideration is required. The description of recent work requires eleven sub-divisions within these three categories, and he presents a chapter on each. The last three of these concern the "response approach" which incorporates and is mainly concerned with the scaling methods attributed to Guttman; and the account of recent developments in the field of latent structure analysis, including some of Dr. Torgerson's own work in the field of multi-dimensional scaling, is likely to be particularly useful in bringing the research worker up-to-date in this difficult branch. It is of interest that the author regards latent structure analysis as "a very general way of conceptualising the measurement problems encountered in the assessment of attitudes, abilities and personality traits."

D. M. EDWARDS PENFOLD.

### PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

- R. M. STOGDILL (1959). *Individual Behavior and Group Achievement*. New York: Oxford University Press, pp. xi+352, 40s.
- S. WISEMAN and J. WRIGLEY (1959). *The Manchester Tests: Mechanical Arithmetic (Senior) 1, Reading Comprehension (Senior) 1, General Ability (Senior) 2*. University of London Press. Arithmetic Test 54s. 3d./100, Reading and General Tests 36s. 2d./100. Manuals 1s. 6d./test.
- F. L. MARCUSE (1959). *Hypnosis: Fact and Fiction*. Penguin Books, pp. 224, 3s. 6d.
- L. M. TERMAN and M. H. ODEN (1959). *The Gifted Group at Mid-Life*. Stanford, Calif.: Stanford University Press; London: Oxford University Press, pp. xv+187, 36s.
- D. G. MORTENSON and A. M. SCHMULLER (1959). *Guidance in Today's Schools*. New York: John Wiley; London: Chapman and Hall, pp. viii+436, 46s.
- G. MANDLER and W. KESSEN (1959). *The Language of Psychology*. New York: John Wiley; London: Chapman and Hall, pp. xviii+301, 54s.
- R. A. HARPER (1959). *Psychoanalysis and Psychotherapy: Thirty-six Systems*. Englewood Cliffs, N.J.: Prentice-Hall, pp. viii+182, \$1.95.
- N. MASOR (1959). *The New Psychiatry*. New York: Philosophical Library, pp. viii+154, \$3.75.
- O. K. BÜROS, edit. (1959). *The Fifth Mental Measurements Yearbook*. Highland Park, New Jersey: Gryphon Press, pp. xxvii+1292, \$22.50.

- E. WOOD (1959). *Yoga*. Penguin Books, pp. 272, 3s. 6d.
- W. DIXON (1959). *Education in Denmark*. London: Harrap, pp. 233+VI, 25s.
- R. COUSINET (1959). *Pédagogie de L'Apprentissage*. Paris: Presses Universitaires de France, pp. 137, NF 4.80.
- S. J. BECK and H. B. MOLISH (1959). *Reflexes to Intelligence*. Glencoe, Ill.: The Free Press, pp. xiv+669, \$8.50.
- C. W. CHURCHMAN and P. RATOOSH (edit). (1959). *Measurement Definition and Theories*. New York: John Wiley; London: Chapman and Hall, pp. viii+274, 64s.
- R. L. THORNDIKE and E. HAGEN (1959). *10,000 Careers*. New York: John Wiley; London: Chapman and Hall, pp. x+346, 68s.
- E. FRASER (1959). *Home Environment and the School*. Scottish Council for Research in Education, Publ. No. 43. University of London Press, pp. vii+83, 10s. 6d.
- A. W. ROWE (1959). *The Education of the Average Child*. Harrap, pp. 255, 15s.
- NATIONAL ASSOCIATION FOR MENTAL HEALTH (1959). *Truancy—or School Phobia*. Proceedings of 15th Inter-Clinic Conference, pp. 40, 3s. 6d.
- I. OPIE and P. OPIE (1959). *The Lore and Language of School-children*. Oxford University Press, pp. xix+417, 35s.
- E. JONES (1959). *Free Associations: Memories of a Psycho-analyst*. London: Hogarth Press, pp. 264, 25s.
- N. CLARKE (1959). *Physics as a Career*. (2nd edit.) Chapman and Hall, pp. 79, 8s. 6d.
- B. BOARD (1959). *The Effect of Technological Progress on Education: A Classified Bibliography*. Institution of Production Engineers, pp. 141, 14s. 6d.
- C. G. JUNG (1959). *The Archetypes and the Collective Unconscious*. Collected Works, Vol. IX, Part. I. London: Kegan Paul, Routledge and Kegan Paul, pp. ix+462, 52s. 6d.
- R. M. GAGNE and E. A. FLEISHMAN (1959). *Psychology and Human Performance*. New York: Holt, pp. xv+493, \$7.25.

# **OCCUPATIONAL PSYCHOLOGY**

*Editor : ALEC RODGER*

JANUARY, 1960

Vol. 34, No. 1

The Sense of Responsibility among Young Workers: Part I, Definition and Measurement.  
By ALAN W. BROWN and HENRY A. LANDSBERGER.

Psychological Factors causing Labour Turnover among Underground Workers. By SAULI  
HÄKKINEN and YRJÖ TOIVAINEN.

The Selection of Nurses: A New Approach. By LIONEL R. C. HAWARD.

Communicating the Findings of Psychological Research in Industry. By T. M. HIGHAM.

Some Factors Affecting Technical Progress in the Cutlery Industry. By H. C. BAKER and  
S. MITCHELL.

15 Plus: Another Problem. By ALEC RODGER.

Book Reviews.

Other Books Received.

*Annual Subscription 40 shillings.*

**National Institute of Industrial Psychology**

**14, Welbeck Street, London, W.1.**



INTERACTION OF HEREDITY AND ENVIRONMENT IN  
REGARD TO 'MEASURED INTELLIGENCE'

D. H. STOTT

*(Department of Psychology, University of Glasgow.)*

SUMMARY. Recent work in experimental biology suggests interaction in the form of the facilitation of genetic pre-disposition by adverse conditions of gestation. This may explain not only how certain environmental factors may produce mental defect and physical malformation in human beings, but also certain more general findings of the Scottish Mental Survey.

The greater resemblance in 'measured intelligence' of monozygotic twins can no longer confidently be attributed to 'heredity.'

Interaction of this type presupposes a very great number of alternative modes of development. Congenital variations in temperament and motivation, by affecting the acquisition of mental skills, add further alternatives.

The 'contributions' of heredity and environment can, consequently, be the subject only of retrospective estimate in individual cases; general quantitative estimates for a population are meaningless.

## I.—INTRODUCTION.

DISCUSSION of the relative contributions of heredity and environment to 'measured intelligence' seems to have reached a point where the value of the dichotomy has been called in question owing to the demonstration of a certain degree of correlation. This correlation, it was pointed out (Maddox, 1957), is due to the sociological fact that high test-intelligence of the parents is associated with high cultural stimulation of the child. Maddox also suggested, from the analogy of chicks, that the manifestation of a hereditary characteristic might be dependent upon post-natal environment.

The object of the present contribution is to draw attention to a body of comparatively recent work in experimental biology which has demonstrated yet another and, indeed, far closer type of interaction, further complicating the heredity-environment dichotomy.

## II.—CONGENITAL MALFORMATIONS IN ANIMALS AND MAN.

It has been known for a generation that the subjection of the pregnant female to certain exogenous stresses was liable to produce malformations in the young. But these were held to be 'phenocopies,'\* resulting from an arrest of embryonic development similar to that produced by a genetic anomaly (Goldschmidt, 1938). The exogenous factors were thought of as alternative to the genetic, and hence the unconvincing attempts to classify individual malformations as either environmentally or as genetically induced. Among the supposed latter sort there was a notable failure to establish a Mendelian pattern of inheritance. Some were attributed to dominant, some to recessive genes, and some to both (Warkany 1947, Penrose 1951); but whichever pattern was held to operate did not do so consistently. However, this variable 'expressivity' of gene influence was a common phenomenon in genetics, and was

\* 'Phenocopy' is the term used for a physical anomaly appearing in a racial strain without any forerunner such as would suggest some form of inheritance—but which is, nevertheless, found to have genetic forerunners in other strains. It is thus said to be a 'copy' of an inherited anomaly.

accounted for by two special explanations—that of the ‘phenocopy’ mentioned above, and ‘penetrance,’ the latter being essentially no more than the statement that an apparently gene-determined tendency achieved morphological manifestation only in a proportion of cases. These terms bear more than a trace of ‘explanation by label.’ Moreover, the ‘copying’ implied in the phenocopy not only implies an agent, but perpetuates the concept of an agent *distinct from* the gene; while the concept of penetrating some unspecified barrier goes beyond methodological parsimony. Both terms have been rendered superfluous by evidence that genetic and environmental factors may interact to produce the malformations formerly attributed to the one or the other.

A wide range of unfavourable environmental factors during pregnancy can produce malformation in the offspring. The type of stress applied—chemical, vitamin deficiency, mechanical, irradiation—is immaterial so far as the type of malformation is concerned; the latter seems to depend upon the stage of gestation, and as mentioned below, upon dormant hereditary propensity. Whereas species of animals vary somewhat in their susceptibility to different stresses, the same general tendency is found in all. For example, vitamin deficiencies have been made to produce congenital malformation in the chicken, turkey, duck, pig, cow, rat, mouse and guinea-pig (Giroud, 1954). No verification by controlled experiment is, of course, possible in man, and except for German measles there is as yet no agreement as to what are the noxious agents. Mumps is under suspicion and so is influenza, but apart from the recent epidemic there has been no good opportunity to enable the guilt of the latter to be tested. In fact, attempts to relate malformation to physical illnesses during pregnancy have been surprisingly unrewarding. Recently, psychological factors have been adduced (Ingalls, 1947, 1956, Stott, 1957, 1958). The sharp rise in the incidence of congenital malformations from the beginning of the war and in the early post-war years in Germany are attributed by German workers in part to emotional shocks and stresses. (Gesenius 1952-2, Klotz 1952, Grebe 1953, Grebe and Windorfer 1953).

Despite disagreement as to which types of maternal stress are responsible, all authorities are agreed that the pre-natal environment plays a large part in the production of human congenital anomalies. It is hard to account otherwise, for example, for the enhanced risk of mongoloid defect and of other malformations with both advancing maternal age, and also in first births (Murphy 1947 Record 1956, MacMahon and McKeown 1953, Alwyn Smith and Record 1955), for the higher incidence of anencephaly (lack of cerebral cortex) among first-births conceived during the summer months (McKeown and Record 1951), or for the association between mental defect and conception during hot summers (Knobloch). The terms ‘congenital’ and ‘innate’ have consequently to embrace pre-natal environmental influence as well as heredity.

It was during the experimental work with animals that the more specific evidence of genetic-environmental interaction was forthcoming. Observing the incidence of diaphragmatic hernia in rats, Anderson (1949) observed that, in a strain which does not spontaneously show hernia, a diet deficient in vitamin A could produce only 9.9 per cent., but in another strain, which showed 2.7 per cent. hernias even with a diet rich in vitamin A, the deficiency produced 19 per cent. Similarly, Fraser and his colleagues (1951, 1954), inducing cleft palate in mice, found that the incidence varied from 4 per cent. to 100 per cent., according to the genotype of the mother and the foetus. Thus, the impairment would seem to result from a hereditary pre-disposition which was *facilitated*, that is, brought into morphological manifestation, by environmental stress during gestation. Facilitation constituted a unitary theory which explained both the undoubted

tendency of certain malformations to run in families and their erratic and also sporadic appearance, without recourse to special explanations such as those of phenocopy or penetrance.

### III.—' INTELLIGENCE ' AND PRE-NATAL ENVIRONMENT.

This new biological concept promises to be relevant to current thought about 'intelligence.' There is a certain amount of evidence (Pasamanick and Lilienfeld 1955, Knobloch and Pasamanick 1956, Stott 1957, 1958) that human mental impairment is related to adverse pre-natal environment. This might well explain why the prediction of a decline in national intelligence was controverted by the Scottish survey. The fall of I.Q. with increase in family size observed therein has its counterpart in the higher incidence of several malformations in later-born children of large families. Anderson, Baird and Thomson (1958) found a correlation between fatal malformation of the central nervous system and adverse social conditions, notably in large cities. If the neural structures upon which mental ability depends are similarly subject to lesser impairment, we should expect a rise in the mean 'measured intelligence' of a population with improvement in social conditions. In effect, mental impairment due to adverse conditions of pregnancy would seem to explain certain trends in the statistics of the Scottish Mental survey that its authors were at a loss to account for. When the fall in test-score with family size is examined closely, it is seen that this was most striking among the 'white collar' groups, and within them at that family size—four or five—at which the latest addition may have been unwanted. The intellectual inferiority of children of mothers of 21 or less is also largely confined to the first four social groups. This can hardly be due to the greater risks of first-births, since the trend virtually disappears in the manual-worker groups. It has been estimated (Wimperis, 1960) from the Registrar General's Statistical Survey that two-thirds of legitimate children born to mothers of under 20, and 24 per cent. of those of 20—24, are pre-maritally conceived; and a significant proportion of the children of the mothers of 21 or under covered by the Scottish Survey would, presumably, be illegitimate. It is reasonable to infer that extra-marital conception would be more frequent among those young mothers belonging to the higher social classes, and the mental stresses associated therewith would be greater among them for cultural reasons. That children of one-child families living in grossly overcrowded conditions should be mentally inferior even to those of larger families in similarly overcrowded conditions again suggests the effects of illegitimacy. There is consequently some reason for inferring that the mental impairment following adverse conditions of pregnancy is not limited merely to that of the occasional production of mental defect (the 'ineducables' were excluded from the survey), but may apply to the merely sub-average.

The hypothesis of the facilitation of genetic tendencies by environmental influences would also explain the failure of mental defect in general, and above all of feeble-mindedness, to conform to genetic expectations (Penrose, 1949). A further important study has recently been published by Brandon (1957), who followed up the children of female mental defectives who had at one time been inmates of the Fountain Hospital, London. Ninety-nine of the 109 children still alive were apparently mentally normal. Only six of the ten subnormals consistently scored below I.Q. 70, and only four below 65. This is higher than would be expected in a general population, but probably about the same as would be found in a slum or other culturally disorganized community. Some three-quarters of the children in question were born out of wedlock and in circumstances unlikely to afford the mother a trouble-free pregnancy or the child an



optimal post-natal environment, so that it is open to conjecture what proportion of the 9 per cent. subnormality was due to heredity.

In a sample of 450 mentally normal children the present writer (Stott, 1957) found that, after stressful pregnancy, one child in three suffered serious non-epidemic illness before the age of 3 years (compared with one in nine after a reported clear pregnancy). The stresses in question were similar in type, although probably on the whole somewhat less acute, than those followed by mental subnormality and/or malformation. The risk of either of the latter types of impairment following a stressful pregnancy can be roughly estimated on the basis of the incidences quoted at between 1 : 30 and 1 : 40. This study must rank only as an exploratory one, but the findings are consistent with the theories of facilitation, and help to explain why mental subnormality is more prevalent in poor neighbourhoods. These findings add infantile ill-health to the 'continuum of reproductive casualty' hypothesised by Warkany, Pasamanick and Lilienfeld (1955), Warkany (1947) and others.

#### IV.—TWINS.

Facilitation has a quite particular bearing upon the twin-studies which are currently regarded as a means of isolating the genetic component of 'measured intelligence.' Maddox (1957) has already suggested that identical twins, even though separated from near birth, will probably have more similar environments—by way of placement with relatives or in selected foster or adoptive homes, etc.—than a random sample of children, so that the whole of the correlation cannot be attributed to heredity. To this we must now add a sharing of the foetal environment, for whatever stresses the mother was subjected to would apply to both. Furthermore, twinning in itself presupposes heightened maternal mortality of twins. If, therefore, pre-natal environment can exercise a general influence upon mental development apart from the production of definite defect, this will also tend to contribute to the correlation of measured intelligence in separated twins.

The Scottish Survey (1953) found that the mean I.Q. of twins was 4.83 points lower than for all non-twins (5.55 points lower for boy-twins). The inferiority of boys to girls among twins (3.06 points compared with 1.77 for non-twins), which the authors of the report comment upon with surprise, is in conformity with the greater general vulnerability of male infants to pre-natal impairment (Stott, 1957). The intellectual inferiority of twins is borne out by Sandon's (1957) study of the relative proportions of twins and of non-twins who get through the selection examination for Birmingham grammar schools. Of the twenty-four twins per 1,000 live births, he estimates that sixteen reach examination age. But only eight twins could be discovered in every 1,000 of the them, but he also found the examination marks of twins to be significantly lower than those of non-twins, and it was probable that the deficiency was more pronounced in the case of the monozygous.

It is possible in short, that the observed higher correlation of 'measured intelligence' between monozygous twins compared with heterozygous twins and non-twins may be accounted for, at least in part, by the theory of facilitation. Given a pregnancy-stress, impairment depends in the first place upon the genetic vulnerability of the foetus to the impairment in question. It would also seem to depend upon the genetic proneness of the mother to withstand stress, and presumably upon her acquired vulnerability thereto. Thus, in monozygous

twins, not only the genetic constitution but also—subject to the qualification below—the foetal environment is similar; for heterozygous twins, the genetic constitution differs but the foetal environment is by-and-large similar; and of course, for non-twin sibs the genetic constitution again differs, while the pre-natal environment (same mother, different pregnancy) will vary more than that for fraternal twins (same mother, same pregnancy). The above is in conformity with the hierarchy of correlations found for children of the above groups reared apart. In addition, monozygous fetuses share the same membrane, whereas the heterozygous do not. Hence, one would suppose that for monozygous twins the pre-natal environment would be somewhat more similar than for fraternal twins. It is again interesting that the Scottish survey found that like-sex twins, both boys and girls, were inferior to unlike-sex twins. The difference is slight, but if it represents an inferiority of the monozygous, as Sandon suggests from a similar result in his own study, it must be multiplied by two, since these would only constitute about half of the like-sex twins. In summary, there would seem to be environmental factors making twins as a whole more similar in mentality than non-twin sibs, and monozygous twins more similar than heterozygous, however soon after birth the twin pairs are separated. It seems thus unfortunately not to be true, as Warburton (1958) avers, that in 'identical' twins Nature has provided us with the means to measure the genetic component.

#### V.—THE INDIVIDUAL'S ALTERNATIVES OF MENTAL DEVELOPMENT.

The value of the hypothesis of facilitation is that it suggests a process by which the interaction of heredity and environment can affect mental development. If it may be assumed that a significant part of the genetic influence is of the facilitateable sort, it is possible to envisage broadly the range of alternatives of mental development open to an individual embryo. At each critical stage of embryonic life, there will be different, genetically determined, pronenesses to impairment. Whether any of these are triggered off, to produce the impairment characteristic of the genotype, depends upon the intra-uterine environment at that particular stage. It is also realistic to think of impairment, not as a lowering of a general 'level of intelligence,' but as damage to the neural structures which make mental operations of certain sorts impossible. This does not necessarily mean that certain *tasks* are made impossible, since we know that similar tasks are often performed by different individuals by very different mental processes. Hence, the alternative process upon which an impaired individual falls back may be just as good, or nearly so, as the primary one. On the other hand, impairment in the form of an inability to form general concepts, or—as suggested below—of intrinsic motivation, may have no substitute and so result in poor general mental functioning. Whether the individual can, in fact, fall back upon alternative processes depends upon whether the environment in which he grows up allows him to exploit the next most suitable alternative, or the next after that, as the case may be. The developmental potential of any individual can thus be likened to the path of a traveller who runs the risk of having to take diversions, each of which offer a variety of obstacles; his ability to surmount any one becomes important only if he is forced to take the path upon which it lies. If he suffers from congenital night-blindness, and gets drowned by falling into a deep pool, his demise might be attributed mainly to a hereditary factor; but if he had been killed by treading upon an unexploded bomb in taking a short cut despite warning notices his death could have been said to be due mainly to environmental factors in the form of defects in his education or upbringing. If one likes to conceive of the roles of heredity and

environment in terms of quantitative contributions, it is consequently seen that the (pseudo-) ratio heredity/environment not only varies from individual to individual, but according to the turn which his development has taken. The value of a general estimate of the ratio, when the variance must be so large, surely becomes meaningless.

#### VI.—MOTIVATION AS A DETERMINANT OF 'INTELLIGENCE.'

Nor, unfortunately, have we exhausted all the complications. 'Measured intelligence' is a sample of an individual's behaviour. Being such, it depends not only upon the possession of an aggregate of neural structures such as make certain mental operations possible, but upon the extent to which the individual is actually motivated to perform them. This is not simply a question of how the subject feels at the time towards the test and the tester. For the whole of his life he will have been variously motivated to think thoughts of a certain type; he will have acquired particular readinesses or mental sets, and the general type and level of his concept-formation—the systems by which he has coded information—will depend cumulatively upon his congenital plus environmentally conditioned motivation.

Now, when we study the behaviour of mentally subnormal individuals without pre-conceptions about a self-operating quantity of 'intelligence,' we observe that impairment of motivation, alias 'emotional factors' (Clarke and Clarke, 1958), is a large element in their poor performance. In some, one suspects, failure or distortion of motivation (in very unresponsive or hyperactive defectives) could be the only factor. In such children nutritional motivation may evoke good perceptual discrimination and appropriate response, but mental activity vanishes when it is no longer a question of eating or drinking; it is as if they had suffered congenital prefrontal lobotomy (Brazier 1950, Lashley 1935). Other defective children, although active, have no interest in human relationships, and never learn speech or social adaptation. A good proportion of any group of educationally subnormal children suffer from 'unforthcoming' temperament, *alias* under-motivation. Other things being equal, this would make for poor concept-formation, since the child would not have performed the mental operations upon which conceptualization is based. In Piaget's (1951) terminology, there would be a shortage of the anticipatory schemata which are built upon previously existing classifications and other learned relations. Our general failure to appreciate the importance of motivation in this building up of the mental equipment is due to the conventional separation of the cognitive and conative fields.

Impairment of motivation—whether it be 'unforthcomingness,' disorganization (inability to concentrate), lack of response to personal stimuli or other abnormality—must therefore be reckoned a component of 'measured intelligence.' In so far as it is congenital, there is some evidence that it can be facilitated by pregnancy-stress in the same way as can cognitive disability (Stott, 1959). This adds a third dimension of alternatives to the possibilities of an individual's mental development.

#### VII.—IS 'INTELLIGENCE' SUBJECT TO ALLOCATION?

At this point it may be well, in summarizing the argument, to see what remains of the concept of relative contribution when it is translated into the language of interaction. The contemporary view of genetic determinants is that they are built-in 'instructions' for the development of the organism within the range of environment for which it is fitted. To ask whether given individual



variations are more due to the variability of the instructions or to that of the environment is not to ask one question, but a thousand or a hundred-thousand—according to where we stop following the alternatives. It would, no doubt, be feasible, in dealing with an individual at a given stage of mental development, to study retrospectively which of the alternatives of interaction were in fact followed, and so be able to say, *of that individual* that heredity or environment played the major part. For example, it might be said of a child of an academic family that his mental development was greatly stimulated by his environment, and that if he had been born in an unacademic family he would have remained intellectually undistinguished; or of a child suffering from a congenital mental impairment that no matter how good the environment his mental development would always have remained low. But to try to assess either of the two types of influence as percentages, even in individual cases, would be merely rhetorical quantification.

Moreover, we are driven to the point of recognizing that we are not asking an answerable question. If an organism failing to solve a problem by one sort of mental behaviour is free within certain limits to utilise other skills; if the relevance of each of the alternative skills depends upon the type of problem; if the primary mode of solution, or the calling into play of the alternative approaches, depends upon the strength and quality of the individual's motivation; if the skills themselves owe their existence in large part to the motivation which has instigated the skill-forming practice—then we are forced to question the value of subsuming the resultant individual differences in mental behaviour within a concept of 'intelligence.' Our ability to measure specified individual differences does not justify us in assuming that there is any unitary factor underlying them, which can be conceived of as a quantity. In our term 'intelligence' we have in fact taken over the popular and traditional reification of these individual differences, which exists not in the mind of the subject, but in that of the observer. Theoretically speaking, the term 'measured intelligence' does not save us, for the attempted measurement of an animistic concept no more conjures it into reality than two wrongs logically ever make a right. By extension, it is also impermissible to pose queries as to the relative quantitative contributions of heredity and environment to mental efficiency unless we specify efficiency in what and in which person, and even then the question only makes sense in broad clinical assessment of etiology.

#### VIII.—REFERENCES.

- ANDERSEN, D. H. (1949). Effect of diet during pregnancy upon the incidence of congenital hereditary diaphragmatic hernia in the rat. *Amer. J. Path.*, 25, 163-84.
- ANDERSON, W. J. R., BAIRD, D., THOMSON, A. M. (1958). Epidemiology of stillbirths and infant deaths due to congenital malformation. *Lancet*. 1958, i, 1304.
- BRANDON, M. W. G. (1957). Intellectual and social status of children of mental defectives. *J. Ment. Sci.*, 103, No. 433, 710.
- BRAZIER, M. A. B. (1950). Neural nets and integration of behaviour. Richner, D. (ed.). *Perspectives in Neuropsychiatry*. London.
- CLARKE, A. D. B., and CLARKE, A. M. (1958). *Mental Deficiency: the Changing Outlook*. London: Methuen.
- FRASER, F. C., et al. (1954). Experimental production of cleft palate with cortisone and hormones. *J. cell. comp. Physiol.*, 43, suppl. 1, 237.
- FRASER, F. C., and FAINSTAT, T. D. (1951). Production of congenital malformation in the offspring of pregnant mice treated with cortisone. *Pediatrics*, 8, 527.
- GESENIUS, H. (1852-2). Genwirkungen und Umwelteinflüsse. *Anat. Anz.*, 98, 272.
- GIROUD, A. (1954). Malformations embryonnaires d'origine carientelle. *Biol. Rev.*, 29, 220.
- GREBE, H. (1935). Zur Ätiologie letaler Missbildungen. *Verh. dtsh. Ges. Path.*, 36, 395-403.

- GREBE, H., and WINDORFER, A. (1953). Beitrag zur erblichen und nichterblichen Missbildungsätiologie. *Dtsch. Med. Wschr.*, 78, 149-151.
- GREGG, N. MCA. (1941). Congenital cataract following German measles in the mother. *Trans. Ophthal. Soc. Austral.*, 3, 35.
- IGALLS, T. H. (1947). Etiology of mongolism. *Amer. J. Dis. Child*, 74, 147.
- INGALLS, T. H. (1956). Causes and prevention of developmental defects. *J. Amer. Med. Ass.*, 151, 1047.
- KLOTZ, R. (1952). Das psychische Trauma in der Genese der Missgeburt. *Z. Gynack.*, 74, 906.
- KNOBLOCH, H. Seasonal variation in the births of the mentally deficient. (Ohio State Univ. paper pending publication.)
- KNOBLOCH, H., and PASAMANICK, B. (1956). Distribution of intellectual potential in an infant population. (Paper presented to conference, Epidemiology of Mental Disorder, N.Y.).
- LASHLEY, K. S. (1935). Studies of cerebral function XI—The behaviour of the rat in match-box situations. *Comp. Psychol. Monogr.*, 11, No. 2.
- LILIENFELD, A. M., and PARKHURST, E. (1951). Association of factors of pregnancy and parturition with the development of cerebral palsy. *Amer. J. Hyg.*, 53, 262.
- MCKEOWN, T., RECORD, R. G. (1951). Seasonal incidence of congenital malformation of the central nervous system. *Lancet*, Jan., 27, 192.
- MACMAHON, B., MCKEOWN, T. (1953). Incidence of harelip and cleft palate related to birth rank and maternal age. *Amer. J. Hum. Gen.*, 5, No. 2.
- MADDOX, H. (1957). Nature-nurture balance sheets. *Brit. J. Educ. Psychol.*, 27, 166.
- MURPHY, D. P. (1947). *Congenital Malformations*. Philadelphia.
- PASAMANICK, B., and LILIENFELD, A. M. (1955). Association of material and fetal factors with development of mental deficiency. *J. Amer. Med. Ass.*, 159, 155.
- PENROSE, L. S. (1949). *The Biology of Mental Defect*. London: Sidgwick and Jackson.
- PENROSE, L. S. (1951). Heredity and environment in the causation of foetal malformations. *Practitioner*, 166, 429.
- PIAGET, J. (1951). *The Psychology of Intelligence*. London, pp. 38-9.
- RECORD, R. G. (1956). Observations related to the aetiology of placenta praevia with special reference to the influence of age and parity. *Brit. J. Prev. Soc. Med.*, 10, No. 1.
- SANDON, F. (1957). Relative numbers and abilities of some ten-year-old twins. *J. Roy. Stat. Soc.*, 120, IV, 440.
- SMITH, ALWYN (1955). A note on mongolism in twins. *Brit. J. Prev. Soc. Med.* 9, 212.
- SMITH, ALWYN and RECORD, R. G. (1955). Maternal age and birth rank in the aetiology of mongolism. *Brit. J. Prev. Soc. Med.*, 9, 51.
- STOTT, D. H. (1957). Mental and physical handicaps in the child following a disturbed pregnancy. *Lancet*, 171, i, 1006.
- STOTT, D. H. (1958). Some psychosomatic aspects of casualty in reproduction. *J. Psychosomat. Res.*, 3, 42.
- STOTT, D. H. (1959). Evidence for pre-natal impairment of temperament in mentally retarded children. *Vita Humana*—2, 125-48.
- WARBURTON, F. W. (1958). Review of 'Social Class and Educational Opportunity' by Floud, Halsey and Martin. *Brit. J. Educ. Psychol.*, 28, 89.
- WARKANY, J. (1947). Etiology of congenital malformations. *Advanc. Paediat.*, 2, 1.
- WIMPERIS, V. (1960). *The Unmarried Mother and her Child*. London: Allen and Unwin.

(Manuscript received, December 20th, 1958.)

# THE TEACHING OF PSYCHOLOGY IN TRAINING COLLEGES

BY K. G. COLLIER  
(*Bede College, Durham.*)

**SUMMARY.** The major function of a psychology course for student-teachers is to cultivate the student's insight into the thoughts and motives of ordinary children. A brief account is given of various methods used at a particular college for this purpose. Reference is also made to the problem of ensuring an effective discipline of students' thinking on the academic content of the course.

## I.—INTRODUCTION.

IN 1960, a three-year course of professional education for teachers comes into existence in England and Wales. The prospect provides a stimulus for the reconsideration of purposes and methods in the teaching of psychology in this setting. There appears to be little previous work available on the teaching of psychology in general,<sup>1</sup> for for student-teachers, or in particular for non-graduate teachers. Oeser, in a stimulating article,<sup>2</sup> argues that psychology should be taught by methods which themselves illuminate psychological principles: for example, the students' experience of working in small syndicates enables them to appreciate the social facilitation of academic learning. In the clinical field Millar and Balint, in two perceptive papers,<sup>3</sup> advocate a close inter-weaving of theoretical study and practical acquaintance with patients. A recent monograph from Keele<sup>4</sup> discusses in an enlightened manner some problems and methods in the teaching of the psychology of personality development to teachers and social workers. Valentine's small book<sup>5</sup> has served for many years as a basis for laboratory work to complement formal lectures and discussions; its orientation, however, is towards a somewhat academic study of principles of learning. In any case a number of new techniques have become available since it was written. The above sources and some others<sup>6</sup> have been considered in working out the methods described below.

It is as well to recognise, as a starting point, that the word 'psychology' has itself a great variety of connotations in the educational world, some gross misuses of the word probably being a useful guide to the prejudices of the students. A student of mine, for example, once said to me, "I get sick of all this psychology in the classroom; why can't you just give a boy a clout or a

<sup>1</sup> But see D. WOLFE, *et al.*, *Improving Undergraduate Instruction in Psychology*. New York: Macmillan, 1952.

<sup>2</sup> O. E. OESER (1951). Psychology applied to the teaching of psychology. *Universities Quarterly*, Vol. 5.

<sup>3</sup> W. M. MILLAR, M. BALINT and G. P. MEREDITH (1954). Symposium on 'Method and Technique in the teaching of Medical Psychology.' *Brit. J. Med. Psychol.*, 27, 30-47.

<sup>4</sup> The problems arising from the teaching of personality development. *Sociological Review Monograph No. 1* (1958).

<sup>5</sup> C. W. VALENTINE (1939). *An Introduction to Experimental Psychology*. University Tutorial Press, 3rd ed.

<sup>6</sup> M. L. JOHNSON, (1952). Teaching by free group discussions. *Universities Quarterly*, Vol. 6.

D. R. MACCALMAN (1953). Observations on the teaching of the principles of mental health to medical students. *Brit. J. Med. Psychol.*, 26, 140-151.



telling off when he needs it and get on with the job?" To him psychology meant never giving direct orders or punishments but getting children to do what one wants by subtlety. I have heard teachers speak as if psychology meant on the one hand intelligence testing and on the other Freud's theories of sex. Such notions represent in my view serious misconceptions of the bearing of psychology on education: fundamentally it should be concerned with understanding people, ordinary people, the ordinary children in the classroom and the playground. To me, psychology for the teacher means primarily developing that immediate intuitive insight into how a boy or girl is thinking or feeling, without which no good teaching is possible.

Several distinct elements contribute to such insight. One needs to be able to understand children's difficulties in academic learning—the mistakes they make in the everyday routine of the classroom—and for that one needs two things. On the one hand one needs that sixth sense a good teacher acquires for nosing out the particular muddle or misconception a child has fallen into. On the other hand, it entails one's being able to estimate the various possible causes of the child's general weakness, which have been investigated by systematic scientific research, such as low general ability, or emotional barriers arising at home or at school, or confused earlier teaching, or an unsuitable syllabus, or sheer lack of adequate stimulus to the child's imagination and interest. One needs to be able to use the standard techniques for checking one's surmises—standardized tests of ability and attainment, inquiry into home background and into relationships with other children and teachers in the school, and so on. One needs to understand children's reactions, both superficial and deep-seated, individual and collective, to different types of discipline; whether—for example—quietness in a particular class means a co-operative attitude, or submerged resentment, or sheer cowed submission to the teacher's power. One needs to appreciate the responses to different kinds of leadership. One needs furthermore to be able to see the long-term influences of one's relations with individuals and classes, and of the general life of the school. Much of one's insight into all these things depends on one's self-knowledge, on one's appreciation of one's own make-up and development, and especially on one's attitude to one's own weaknesses and childish failings. A teacher who always treats untidiness with severity has failed to realize that his own latent tendencies to untidiness are probably held in check by a somewhat authoritarian conscience or super-ego.

A major problem for the education lecturer in a training college, then, is how to cultivate such insight in his students, how to open their eyes to what is happening in the children in front of them. Some people think it cannot be done at all; one has much or little of the gift by nature and it cannot be improved by training. Others think it can be taught, and that lectures and discussions are an effective instrument. My own opinion is that it can be improved by tutors but that lectures and discussions cannot in themselves do much. There are doubtless a great many different methods and the following is an account of those used in one college.

## II.—METHODS USED FOR DEVELOPING STUDENTS' INSIGHT INTO CHILDREN.

(1) *Pre-entry observation.* We arrange for students to spend a fortnight in schools for observation before entry to the college. We supply questionnaires on School Organization and on Children's Behaviour, and it is noticeable that the students find the latter far more difficult than the former. Even a rather detailed series of questions means little, apparently, to an adult of 20 who has no standard by which to judge—for example—whether children's groups are

'stable' or 'unstable,' or how far a child is 'dependent' on the approval of adults.

(2) *Films.* We use a variety of films, from the Ministry's special training college films to a Russian account of conditioned reflexes and Slavson's record of the group-treatment of maladjusted ten-year-old boys.<sup>7</sup> Such films provide excellent concrete experience to be shared by tutors and students; they need to be prefaced by an introduction and to be followed up with discussions in small groups. Some of the most profitable films we have come across are a couple of U.P.A. cartoons.<sup>8</sup> The length is only ten minutes, the drawing and scripts are delightfully witty, and psychological principles are faithfully illustrated. One can easily repeat the showing after discussion.

(3) *The tape-recorder* is a useful instrument. One can record children talking together and discuss the interpretation of the conversation. One can select fragments from recorded lessons to illustrate various methods of exercising authority and play them back in a lecture. One can record a student's lesson and discuss both his strong and his weak points and their origin and significance for him as a teacher. This is best done in a tutorial group where the man's colleagues can help him—in the right atmosphere—to look at himself more candidly. I am sure there are many as yet unexplored uses of the tape.

(4) *Literature.* One can discuss selected films, novels, plays, biographies etc., with a group all members of which have shared the experience. We have used such films as *The Bridge on the River Kwai*, *The Third Man* and *The Grapes of Wrath*; and such novels as *David Copperfield* and *Sons and Lovers*. These all provide excellent opportunity for discussion of the springs of human motive.

(5) *Self-Study.* We also value highly, as a means of encouraging students to embark on the long job of understanding their own make-up, the setting of an essay on their own development immediately after the set lectures and discussions on the development of personality. We ask specifically for a report of 500-1000 words on some of the individual student's own chief attitudes and values, referring particularly to those he has acquired through identifying himself with a loved adult, and those in which he has been drilled by someone in authority. We warn them, of course, that we have no wish to pry into private affairs. Many discover in this exercise a new understanding of life and people. With an adequate tutorial system one could get more from these self-studies than we usually do.

(6) *Learning activities.* The great majority of our students have been through years of dry academic instruction; they have very little experience of any method of study but 'swotting.' They therefore have the greatest difficulty in grasping what one is talking about when one says that historical episodes can be made more vivid to children by dramatization, and yield more insight; or that concepts are useless unless they have an adequate background of concrete experience. We, therefore, give them a series of miniature learning activities. (a) We divide a discussion group into four sub-groups and give them a piece of unfamiliar history to learn by four different methods—mime; mapping; race-games; and swotting. We test them for factual knowledge a few weeks

<sup>7</sup> *Activity Group Therapy*, obtainable from Geneva via British Committee for Interchange of Social Workers and Administrators, 26, Bedford Square, London, W.C.1.

<sup>8</sup> *Family Circus* and *Willie the Kid*, available from Wigmore Films, 142-150 Wardour Street, London, W.1.

later and show that there are negligible differences. Readers will recognize this scheme as based on Northway.<sup>9</sup> (b) We give them a series of ten-minute studies to do in pairs; learning an Ethiopian vocabulary by means of race-games; making comparative observations on specimens of twigs; classifying three dozen pictures post-cards into six categories; drawing conclusions from a series of historical or geographical pictures; finding answers to questions on a play or a chapter of the Bible. (c) We require all students to make a model from scrap materials, on some academic topic and to state what academic matter a child should learn through the process of making the model. Thus, they can begin to appreciate the great variety of approaches to study and learning, and the point of one's general statements on the psychology of learning.

(7) *Role-playing.* We use role-playing, or the acting out of general principles in terms of dramatized episodes, for pedagogical principles rather than psychological, but the two overlap. For example, when I deal with principles of authority in schools, I suggest that if a teacher wishes to establish his pupils' confidence in him he must show that he really cares about individuals in trouble. A small group of men then perform a playlet. In the first scene a conventional type of teacher exhorts a troublesome boy to pull himself together or he will get drastic treatment. The second scene is of the boy arriving home for tea and finding nobody at home but an older brother, who is just going out. Dad is on overtime, Mum has just got a part-time job as a barmaid, there is no tea and no fire, and when the older brother goes out the boy exclaims in bitter despair, "Nobody cares a damn for me." Besides the pedagogical question of the teacher's treatment of the boy, there is the psychological question whether it is generally true to say that a troublesome boy is reacting to bad relationships at home. Here such evidence as that of Bowlby and the Gluecks becomes relevant. In our experience, the role-playing approach makes a sharper impact on the student's imagination than a general discussion.

(8) *Study-practices.* By far the most valuable method we have worked with is a system of group practices which we call study-practices, since they are essentially an aid to study; the study of principles of psychology and of teaching method. Under this system every student in his second year works every week for a double or triple period as one of a group of a dozen or so students under a tutor in a class of children. I have a group doing science in a secondary modern 2A class, and another group studying film and radio with a 3B secondary modern class; we have groups taking secondary modern Geography, English and Mathematics; and groups in junior schools doing Reading and Number Work, Art, Music and Nature Study. There is some class teaching but the greater part of the work is carried out in groups of two men with four or five children. There are weekly tutorials in the college, in which the next programme is planned and results discussed. Men get to know their own special pupils well and, by arrangement with the Head and the parents, visit their homes. In this work we find the ideal opportunity for relating concrete experience and problems to general principles: for example, the characteristics of individual children in relation to those of their age-group; the relation of a child's personality to his home; the responses of children to different teaching methods; the administration and interpretation of standardized tests. Students write a report at the end of the year which is the major written work of the second year and

<sup>9</sup> M. L. NORTHWAY (1940). The concept of the 'schema.' *Brit. J. Psychol.*, 30, 316-325; 31, 22-36.



includes two child studies. At present we get about twenty weeks in the year for study-practice; in the three-year course it will probably be expanded.<sup>10</sup>

### III.—SYSTEMATIC EXPOSITION.

I have described some of the methods by which we attempt to cultivate the student's insight into human motive and thought. They tend to be unsystematic in character, and often to depend on the tutor's spotting the significance of particular incidents, thrown up by chance, for the illumination of general concepts. But we do also make a systematic exposition of theory in mass lectures accompanied by tutorial discussions. We take the usual topics: the stages of development of children; the psychology of learning; the nature of intelligence; the use of psychological tests; the development of personality, including especially the acquisition of attitudes and values; the psychology of authority and leadership; and that of expressive behaviour, such as drama and painting. We shall certainly maintain these in the three-year course. We consider that the element of systematic exposition is important and we are, I think, generally scrupulous in presenting balanced evidence, with statistical corroboration where possible. But we do not give so much time to these aspects, and we refer to scientific method and conceptual definitions in passing rather than in separate sections of the course. In other words we regard our work as a technology rather than as a pure science. Incidentally, we pay considerable attention to our techniques of examination and to elucidating for the students the criteria by which we mark the scripts. We stress the weight given to qualities of insight, judgment and perspective, and to evidence of some interaction between abstract thinking and personal experience in the student's mind, and we are experimenting with questions which will encourage reflection on experience rather than memorization of second-hand ideas.

Our time allocation varies during the two years. In the first year a student has two formal lectures (to a whole year of 200 men) and two discussion periods a week for Education, about a third of the total being given to practical work such as I have described. In the second year Education course a student gets one year-lecture and one discussion linked to it, together with a double period of Study Practice and an associated discussion. A little under half of the second year work is practical.

### IV—SOME DIFFICULTIES.

It will no doubt be remarked that the success of most of this work depends on the size of the tutorial groups. In a recent investigation it was found that in approximately half the colleges studied the ordinary working groups for Education were between twenty-five and forty in size. When I compare these numbers with those in academic subjects where tutorial groups in the same colleges may contain perhaps ten—twenty students, I reach the conclusion that the most important single need for the teaching of psychology in the three-year course is for a reduction in the size of the normal working groups in Education departments to the standard of the academic subjects. This would, incidentally, enable one to set more written work for tutorial discussion.

<sup>10</sup> See L. W. DOWNES, *Bulletin of Education*, May, 1953.

K. G. COLLIER, *Education for Teaching*, May, 1955.

J. J. FIGUEROA, Post-graduate teacher education; some experiments in the U.S., *Universities Quarterly*, Nov. 1958.

P. E. VERNON, Post-graduate training of teachers in psychology. *Brit. J. Educ. Psychol.*, 1950, 20, 149-152.

A second difficulty is getting students to read effectively. An incentive in the shape of a degree or diploma would probably stimulate more reading: the number of books borrowed from the college library always rises before an examination. But it is the *quality* of reading that troubles us: we find little evidence of the student making any strenuous or imaginative effort to interweave his personal experience and his book-reading, except in the context of the Study-Practice. It may be that this kind of reading only begins when a student gets 'bitten' with a subject and that the real problem is to provide conditions in which this is likely to take place. Even then he probably needs training, particularly in view of the scarcity of books written at his level.

A third difficulty in the way of the teaching of psychology is an ethical one. The successful understanding of a problem in physics or geology depends largely on high ability and intense interest. One analyses the phenomenon and speculates on possible interpretations; one then checks one's hypothesis. One can do something similar in diagnosing the causes of a child's backwardness in reading. But when it comes to dealing with problems of behaviour and discipline one is on different ground. A 14-year-old boy, we may imagine, is off-hand in his manner: one teacher sees in this a scarcely disguised impertinence; another a defensiveness springing from uncertainty of his own status; a third a laudable assertion of his independence. And the difficulty of checking the hypothesis derives not merely from the absence of sufficiently delicate tests but, from the fact that the teacher's personal, deeply-rooted values are involved. In the first teacher the interpretation may be due to the suppression of his own rebelliousness by a punitive conscience; in other words he has learned to set a compulsive value on prompt deference to authority. The second for some reason has come to attach a greater importance to the under-dog's point of view; the third attaches a positive value to an individual's self-assertion irrespective of its cause. When we are engaged in cultivating the student's insight into people, we are unavoidably involved in the possibility of altering his personal values and outlook on life. That is a serious undertaking and a considerable responsibility. It demands rather a high degree of personal integrity and detachment. Perhaps after all it is 'safer' to teach psychology through formal lectures and discussions which cannot get under anyone's skin.

*(Manuscript received December 4th, 1958).*

# A STUDY OF THE CONSERVATION OF SUBSTANCE IN THE JUNIOR SCHOOL CHILD

By K. LOVELL AND E. OGILVIE  
(*University of Leeds Institute of Education.*)

**SUMMARY.** Almost all the children in a Junior school were tested individually, along the lines initiated by Piaget, in an effort to trace the development of the concept of invariance of substance, and to establish the arguments used by children to justify their answers. The three stages in the development of the concept given by Piaget have been confirmed, and some children give reasons for conservation similar to those proposed by him. On the other hand, our evidence does not always agree with that of Piaget, nor does it enable us to prove or disprove the assumption that the child arrives at the concept of conservation because he is able to argue logically in concrete situations.

## I.—INTRODUCTION.

THE concept of conservation or invariance of substance is an important one. A quantity such as a lump of plasticine, a collection of beads, a length or a volume can be used by the mind only if it remains permanent in amount and independent of the re-arrangement of its individual parts. This notion of invariance is indeed essential to any kind of measurement in the physical world. Piaget, using many ingenious experiments, set out to determine if the concept is there in the child's mind from the start or whether it is gradually built up.

It is now becoming generally known that, in Piaget's view, a child's thinking is largely dependent upon perception from 4-7 years of age. During this period thinking tends to be determined by the child 'centring' on one aspect, dimension or element of the situation with other aspects ignored. But, from 7-8 years of age he is able to break away from the influence of perception and is increasingly able to apply logical thought to practical problems and concrete situations. Piaget maintains that the concepts that figure in logical thought result from the co-ordination of actions in which the child combines, dissociates, orders and sets up correspondences. The child gradually appreciates the significance of his actions so that they acquire the form of reversible operations in the mind and thus render his thinking less dependent upon perception.

From his many experiments involving the conservation of continuous and discontinuous quantities, Piaget maintained that children passed through three stages; namely, that of non-conservation, transition and conservation. In a typical experiment, which we have repeated, Piaget or his students would show a child two balls of modelling clay of the same size, one of which is then shaped to look like a sausage. At the first stage the child will deny that the amounts of clay in the 'sausage' and ball are now the same. A response such as "Its more because the sausage is longer," might well be found. At the transition stage, he will arrive at the idea of conservation under some conditions, or at one moment, but will lose the idea again under slightly changed conditions; while from 7-8 years onwards it is said that the child feels the logical necessity for conservation and will support it by argument. For example, Piaget (1950, p. 140 and 1953, p. 16) maintains that the child will say that the sausage can be returned to the shape of a ball, or what has been lost in one direction has been gained in another, or that nothing has been added and nothing has been taken away.



However, we cannot say how many children were examined by Piaget and his students in this type of experiment, and it was decided to test a group of English children under somewhat more controlled conditions.

## II.—EXPERIMENTAL PROCEDURE.

This study reports one main and two subsidiary experiments, the latter being of use in clarifying certain points. In the main experiment almost every boy and girl in a junior school in a North of England town was tested individually in an effort to trace the development of the concept of invariance of substance, and to establish the arguments children use to justify their answers. Balls of plasticine were used of diameter about two inches. Providing the child agreed that the amounts of plasticine in the balls were the same to begin with, the subject was considered to be suitable for the experiment. One ball was then rolled out to make a 'sausage' and at the same time the child was questioned. All the experiments took place in the Headmaster's room so that no child witnessed the cross examination of another, and the children's replies were recorded verbatim. The procedure which was followed is now given, although it must be stressed that many supplementary questions were asked, and our outline indicates the general form of the procedure and not its exact details. To get the child to give his reasons, we found it necessary to make our approach a flexible one.

**APPARATUS.**—Six balls of plasticine, two equal in size, the others clearly different, and all six of different colour.

**TECHNIQUE.**—Ask the child to choose two balls of plasticine which are the same. Say—*Here we have six balls of plasticine. Can you choose the two that are the same?*

If the child replies that they are all different say—*True, in colour they are different but I'm thinking of size, of amounts of plasticine. Let us forget the colours; now, which two are equal?*

Child having made the correct choice, say—*Now you have one ball and let me have the other. Will that be fair?*

(1) *Will I have as much plasticine as you?* The reply must be *Yes* before the experiment continues. Equality having been admitted, say—*I am now going to roll your ball into a sausage.\* Watch me. Action performed. So now you have a sausage and I still have a ball.*

(2) *Who has the most† plasticine now?*

(3) *Why do you think so? What makes you so sure?*

(3) (a) *If the sausage is longer why doesn't it have more plasticine in it? Or, How can it be longer and still the same?* In the case of 'don't know' replies to (3) (a) the child was asked: *What happens as I roll the ball?* Invariably the reply was *It gets longer*. The child was then asked *Anything else?* and, if necessary, *Does it get fatter?*

This was the end of the experiment for those clearly conserving. For non-conservers say—*Watch carefully—I'm rolling your sausage longer still . . . longer still.*

(4) *Now who has most plasticine?*

(5) *Why do you think that?*

(6) *Very well, watch carefully. Let's roll it some more . . . some more. Now who has most plasticine?*

\* Sometimes the child did the rolling but more frequently it was done by the experimenter.

† We have used the word 'most' instead of the words 'the more.' Although grammatically incorrect, we think that the children better understood our question using 'most.'

(7) *Why do you think that?* Point out a second aspect of the dimensions—length or breadth—which the child is ignoring by asking the following kind of question. *But my piece of plasticine is fatter. Does this mean that I have more or less plasticine than you?* Pursue this point with similar type questions. The primary aim here is to force the child to consider two dimensions simultaneously and see what changes in thought occur—if any.

(8) *Let's roll your sausage back into a ball.* Action performed. *Who has most plasticine now?*

(9) *Who will have most plasticine if we roll your ball out again into a sausage?*

(10) *Now watch carefully. I am rolling my ball into a sausage.* Action performed at same time. *Who has most plasticine now?*

(11) *Why do you think so? What makes you so sure?*

The first of the subsidiary experiments was carried out to see if an understanding of addition and subtraction of small pieces of plasticine inevitably produces the concept of conservation. If this is so, then any child who can see that adding / subtracting a piece of plasticine makes a ball more / less, should maintain his conclusion without regard to subsequent spatial transformation. Thus, a piece of plasticine was removed from a ball and the child was then asked what he thought about the amount of plasticine in the ball. The piece removed was left in the clear view of the child. The experimental ball was then subjected to a spatial transformation and the child questioned again about conservation of plasticine. The following numbers of children were involved: thirty-five at the stage of conservation as found from the main experiment, forty-three at the stage of transition, and thirty-eight at the stage of non-conservation, making one-hundred-and-sixteen pupils in all. Some of the children were subjected to this test at the time of the main experiment.

The second subsidiary experiment was designed to test the hypothesis that a child at the stage of non-conservation in the case of a plasticine ball, is not inevitably a non-conserver in another situation involving quantity; in this case a non-conserver of rubber in a rubber band. Thus, a rubber band was shown to the child; it was then stretched, and he was questioned as to whether or not there was the same amount of rubber in the band as before. Forty-eight children all at the stage of non-conservation in the main experiment undertook this test.

### III.—STATEMENT OF THE RESULTS.

#### MAIN EXPERIMENT.

The three stages in the development of the concept of conservation of substance have been found. Table I shows the number of children at each stage in each of the year groups. It was difficult at times to classify correctly a child at the transition stage.

TABLE I  
NUMBERS OF CHILDREN IN EACH GROUP AT VARIOUS STAGES.

	Average Age Yr. Mo.		Conservation	Transition	Non- Conservation	Total
1st year . . . .	7	8	30	27	26	83
2nd year . . .	8	10	44	8	13	65
3rd year . . . .	9	9	73	15	11	99
4th year . . . .	10	8	64	7	4	75
			211	57	54	322

Table 2 shows the percentage number of children at the conservation stage giving various reasons for conserving. Some children gave, spontaneously, more than one reason; and in order to make the meaning of the words used in the table as clear as possible, definitions are first given.

*Reversibility.*—It is difficult to be sure exactly what Piaget means by 'reversibilité.' Bunt (1950, p. 28) thinks he means that the child must be able to picture to himself the 'sausage' returned to the original ball. We were unable to keep to this exact meaning for *reversibility*. Our criterion is that the child gave evidence of his awareness of the equality of the balls at the beginning. The child says in effect, "They were the same before." Of the conservers only five children spoke of rolling the sausage back to a ball which would be identical to the other ball, or pointed out that after rolling back it would be as it was before being rolled out.\*

*Identical action.*—Child says, in effect, that the ball held by the experimenter can be rolled out into a similar sausage.

*Plus/Minus.*—Child says, in effect, that nothing has been 'added to' or 'taken from' when the ball was rolled into a sausage.

*Co-ordination of relations.*—Child says, in effect, that although the sausage is longer it is thinner. Bunt (*op. cit.*) criticises Piaget for the use of the term 'multiplication of relations,' and we have kept to the term 'Co-ordination of relations in compensatory fashion.' It appears that, for Piaget, the word 'multiplication' means what is sometimes called 'logical multiplication,' i.e., two factors simultaneously affecting a third. In this case it is height and cross section jointly affecting volume and he seems to imply something that can be expressed in units. We are not able to keep to Piaget's meaning at this point, if we interpret him correctly.

*Weight.*—Child refers to the fact that the weights of the sausage and ball are the same. The assumption is that the weights of the two balls were the same originally.

*Shape.*—Child says, in effect, that the lump of plasticine has only undergone a change of shape.

TABLE 2  
PERCENTAGE OF CHILDREN IN EACH YEAR GROUP GIVING VARIOUS ANSWERS FOR CONSERVATION.

	Reversibility	Identical Action	Plus/Minus	Co-ordination of relations	Weight	Shape	Don't Know
1st year .....	57	3	30	30	0	7	13
2nd year .....	58	12	42	14	5	2	0
3rd year .....	71	7	29	33	1	0	0
4th year .....	78	5	19	36	6	6	0

We now analyse the replies of the children at the stages of transition and non-conservation. Table 3 shows the number of children in each year group who gave clear evidence of 'centring' on one dimension. The criterion was that they replied, "It's more because it's longer" or some essential variation of this at some point. It must be stressed that others might well have 'centred' on one aspect but we have no clear evidence of it.

\* It will be seen later that some children at the other two stages spoke of rolling the sausage back to a ball which would be bigger/smaller than it was originally.



TABLE 3

NUMBER OF CHILDREN IN EACH YEAR GROUP AT TRANSITION AND NON-CONSERVATION STAGES WHO GAVE EVIDENCE OF 'CENTRING' ON ONE DIMENSION, AND OF A CHANGE OF 'CENTRING.'

	Centring		Change of Centring	
	Transition	Non-Conservation	Transition	Non-conservation
1st year .....	21	23	13	14
2nd year .....	2	3	0	2
3rd year .....	5	10	3	5
4th year .....	6	4	3	3
	34	40	19	24

If a child judges by look alone, the aspect of the total situation on which he is 'centring' may change. The sausage is bigger one moment and the ball is bigger the next. There is no reason why the child should not switch his 'centring,' and if he is not conserving then his opinion might change. The criterion was that the child said that a particular object was, say, 'more' at first and later 'less,' *without* any prompting or forcing by the experimenter;\* that is *before* stage 7 in the general procedure. Again, other children might have changed their 'centring' from one aspect to another but we have no evidence of their so doing. Table 3 also shows the number of children at the transition and non-conservation stages who changed, without any prompting or forcing, the aspect of the situation on which they 'centred.'

At stage 7 in the testing procedure the child was forced to observe two dimensions simultaneously. Table 4 shows the number of children in each year group who give clear evidence of being able to consider two dimensions together. The criterion was that, in essence, the child made reference to two dimensions in compensatory fashion, e.g., "Its longer but its thinner." Table 4 also shows how the children reacted. Those at the transition stage are sub-divided into those who conserved to the end of the experiment and those who conserved for a while but reverted to non-conservation before the end.

TABLE 4

NUMBER OF CHILDREN IN EACH YEAR GROUP AT TRANSITION AND NON-CONSERVATION STAGES WHO GIVE EVIDENCE OF ATTENDING TO TWO DIMENSIONS.

	TRANSITION		NON-CONSERVATION
	Can consider two dimensions and conserve at end of experiment	Can consider two dimensions and temporarily conserve	Can consider two dimensions but fail to conserve
1st year .....	3	3	1
2nd year .....	0	1	3
3rd year .....	0	2	1
4th year .....	1	1	1
	4	7	6

\* It must be realised that the fact that the child undergoes the experiment at all, does to some extent force him to think.

The study of reversibility at the transition and non-conservation stages is of great interest and importance. Table 5 shows the number of children in each age group who gave clear evidence of reversibility as we have defined it. It will be noted that reversibility does not necessarily produce conservation in spite of the fact that it is given as a reason for conservation among conservers.

TABLE 5

NUMBER OF CHILDREN IN EACH AGE GROUP AT TRANSITION AND NON-CONSERVATION STAGES WHO GAVE EVIDENCE OF REVERSIBILITY.

	Transition	Non-Conservation
1st year .....	21	5
2nd year .....	8	10
3rd year .....	11	9
4th year .....	3	3
	43	27

Thus, seventy children can show reversibility as we have defined it, but draw no conclusion from it and remain at the non-conservation or transition stages.

In addition, two children at the transition stage and four at the non-conservation stage showed clear awareness of the possibility of returning the sausage to a ball, but claimed that the ball would be bigger/smaller than the original.

#### SUBSIDIARY EXPERIMENT 1.

A small piece of plasticine was taken from the ball. All the children tested, except two, agreed correctly that there was then less plasticine in the ball. The plasticine was then subjected to a spatial change. The thirty-five children who were conservers in the main experiment remain conservers, while Table 6 shows the effect of the experiment on children at the transition and non-conservation stages.

TABLE 6

NUMBERS OF CHILDREN ATTAINING CONSERVATION AFTER THE SUBTRACTION OF A PIECE OF PLASTICINE.

	TRANSITION		NON-CONSERVATION	
	Examined	Conserved or almost so	Examined	Conserved or almost so
1st year .....	17	10	16	0*
2nd year .....	5	3	10	2
3rd year .....	14	8	8	4
4th year .....	7	3	4	3
	43	24	38	9

\* Two pupils seemed to move some way towards conservation.

## SUBSIDIARY EXPERIMENT 2.

In this experiment we tested to see if non-conservers in the main experiments were also non-conservers in another situation involving invariance of quantity—in this case, of rubber in a stretched rubber band. The results are given in Table. 7.

TABLE 7

NUMBER OF NON-CONSERVERS (MAIN EXPERIMENT) IN EACH AGE GROUP AT VARIOUS STAGES WHEN CONSIDERING THE QUANTITY OF RUBBER IN A RUBBER BAND.

	Number tested	Conservers	Undecided	Non-conservers
1st year .....	26	5	2	19
2nd year .....	11	4	0	7
3rd year .....	7	5	0	2
4th year .....	4	3	1	0
	48	17	3	28

## IV.—DISCUSSION OF RESULTS.

Strong evidence has been produced in support of the three stages proposed by Piaget, and in our view, he was justified in trying to trace the development of the concept of invariance of substance. But the stages are not clear cut; the borders between them are zones not lines. Further, we cannot be sure that the group of operations proposed by Piaget of reversibility, combination of compensated relations, and identity, are, in fact, sufficient, although they may well be necessary before the child can attain conservation.

In the case of non-conservers, there is clear evidence that thinking is frequently dominated by perceptual 'centrings,' usually on single dimensions or equalities. Table 3 shows that forty out of fifty-four children at this stage replied to the effect that "its more because its longer" or some essential variation of this. The remaining fourteen either could not give a reason or advanced all manner of apparently irrelevant arguments such as, "mine's a sausage": "you've rolled it." Furthermore, Table 3 shows that twenty-four non-conservers gave clear evidence of a change of 'centring' during the experiment and it is possible that this double 'centring' may influence the attainment of conservation in some instances.

Piaget has unequivocally stated (1950, p. 140 and elsewhere) that the co-ordination of relations (e.g., height makes up for what is lost in length) is necessary before conservation comes about. From Table 4 we see that six non-conservers can consider two dimensions in compensatory fashion when forced to do so, but fail to conserve. Of the children at the transition stage, seven conserve temporarily when forced to consider two dimensions, while a further four conserved at the end of the experiment.

Again, Piaget has stated (1950, p. 141) that the child at the non-conservation and transition stages had, on occasion, admitted a return to the starting point, without this 'empirical reversal' constituting a complete reversibility. We do not understand exactly the difference between this 'empirical reversibility' and a complete reversibility, but we do know from Table 5 that some children gave evidence of reversibility, as we have defined it, and yet did not conserve.



At the transition stage suggestion is strong in some children. They may give the answer they think is required and not their own opinion, but it was not found difficult to obtain, in answer to later questions, what seemed to us to be their own viewpoint. At this stage the child may begin the experiment by conserving but later resort to non-conservation; or he may begin as a non-conserver, then conserve as the experiment proceeds, and finally revert to non-conservation again. The striking feature about children at this stage is that they are uncertain. The child is stumbling towards the concept of conservation and may revert to non-conservation at any time.

In the final stage the equality of the ball and sausage, from the point of quantity, is always maintained. But one can still note the lingering influence of perception because some children will readily admit that the 'sausage' (or ball) *looks* more. In the first year, 13 per cent. of the age group were unable to give a reason, but all the other pupils did in all the year groups, and many spontaneously gave more than one reason. Some say, "As they were the same before" (reversibility); some say, "You haven't added any," or "You haven't taken any off," (operation of plus/minus); others say, "Although its longer its thinner" (co-ordination of relations).

A small number of children justify conservation on the grounds that the ball held by the experimenter could be changed into a sausage similar to that held by the testee (identical action). A few children say that their lump of plasticine has only undergone a change of shape, or that the weights of the ball and sausage remain constant. Piaget does not mention these reasons as far as is known to the writers. The last is particularly interesting for Piaget maintains that conservation of weight comes later than conservation of substance. He may well be correct and experiments are being carried out to verify his views. But even now we suggest that substance is conserved more safely when supported by conservation of weight.

In the first subsidiary experiment we have shown that witnessing and agreeing to the operation of subtraction of a small piece of plasticine does not in itself necessarily produce conservation. But it is true that the operation does help the child to fix his attention on quantity rather than on dimension, and it is this that seems to help children to conserve who were previously not doing so. About a quarter of the children tested who were non-conservers in the main experiment, now conserve or almost do so. Among those at the transition stage the move towards conservation is roughly doubled, but this could have been partially anticipated since the lines between the stages are not clear cut. It is the decision in the main experiment, to place all those who change their mind into the transition stage, whether they conserved to the end or not, that enables some children to become conservers after this operation with apparent ease.

In the second subsidiary experiment we have demonstrated that children who are non-conservers in one situation involving quantity, are not inevitably non-conservers in another. Indeed, about one-third of those who were non-conservers in the experiment involving plasticine were conservers in the experiment employing the rubber band. Hyde (1959) also found that some children who were non-conservers in a test using plasticine balls were conservers when a liquid was poured from one vessel into another of different shape. For example, when liquids in similar vessels A and B, which certain children agreed were equal in amount, were poured into glasses C and D, respectively, they agreed that the amounts in the latter vessels were equal in spite of great perceptual differences due to C and D being of entirely different shapes. Yet, when the liquids in A and B were matched again for equality, and the liquid in B was poured into a number of smaller vessels which were all similar to one another, some of those who previously agreed to conservation, now denied it. Beard

(1957) also found among sixty children aged 6 and 7 years, that some who were non-conservers when comparing balls of plasticine, could conserve when water was poured from one vessel to a number of smaller ones.

Thus, it seems that children who are conservers of continuous quantities in one situation are not inevitably conservers in another. On this point the more recent findings seem to lead to conclusions at variance with those of Piaget. Our interpretation of his view is that once the concept of the conservation of substance has been attained, it holds in all situations involving conservation of substance. If we have understood him correctly, then our evidence, also that of Hyde and Beard, does not support his viewpoint, which seems to us to hide the particularities of child thinking. It seems rather than the concept is applicable only to highly specific situations at first and that it increases in depth and complexity with experience and maturation. Piaget's view that the child arrives at the concept because he is able to argue logically in concrete situations may or may not be correct. A careful scrutiny of all our evidence does not enable us to prove or disprove his viewpoint. It is equally likely that the concept of conservation of substance—or indeed, any concept—grows out of the interlocking of several organisations of past impressions that normally remain outside consciousness (schemata) which in turn grow out of many and varied experiences. The child may then invoke logical argument to justify the concept which was obtained on other grounds. Or, it may be that experience and the ability to use logical thought aid one another and bring about certainty.

The following points also arise from our study :

(1) Our evidence suggests that in the first two years of junior school there are many children who have not attained much concept of the invariance of substance. It must not be assumed that the children of this school were entirely representative of British junior school children generally, and it is our opinion that in some other junior schools the proportion of conservers might be somewhat higher.

(2) From work reported here and elsewhere, there is, in our view, no doubt that Piaget's type experiments are in themselves learning situations.

(3) Our experiments have shown the grave dangers involved in accepting a child's apparent understanding at face value.

(4) Non-conservation of substance may affect the lives of children and less able adults more than is generally recognised. For example, we have evidence that some children are prepared to pay more money for a piece of toffee when it is in one shape than when it is in another.

(5) Our work has confirmed the considerable verbal confusion known to exist in children in the first two years of the junior school. They confuse such terms as *longer*, *fatter*, *shorter*, *bigger*, *thicker*, *smaller*. Here are two examples from first-year children :

"It gets thinner and bigger, it doesn't get fat though." By *bigger* we assume he means *longer*.

"It'll get right long and smaller and smaller." By *smaller* we assume he means *thinner*.

Piaget, too, has read into the child's remarks what he thought the latter meant, although we cannot be sure that either Piaget or ourselves always judged correctly on this issue. To us this confusion supports our view that the idea of quantity grows slowly.

(6) Reversibility, in the sense in which we have defined it, is the most frequently given reason for conservation at all ages.

## V.—REFERENCES.

- BEARD, R. M. (1957). *An Investigation of Concept Formation Among Infant School Children*. Unpublished Ph.D. Thesis : University of London.
- BUNT, L. N. H. (1950). *The Development of the Ideas of Number and Quantity according to Piaget*. Gronigen, Djakarta : J. B. Walters.
- HYDE, D. M. (1959). *An Investigation of Piaget's Theories of the Development of the Concept of Number*. Unpublished Ph.D. Thesis : University of London.
- PIAGET, J. (1950). *The Psychology of Intelligence*. London : Routledge and Kegan Paul.
- PIAGET, J. (1953). *Logic and Psychology*. Manchester : Manchester University Press.

(Manuscript received, May 29th, 1959.)



# AGGRESSIVE AND WITHDRAWING CHILDREN IN THE NORMAL SCHOOL

## II—DISPARITY IN ATTAINMENT

By E. A. LUNZER

(*Department of Education, University of Manchester.*)

**SUMMARY.** In a recent article, Lynn put forward the view that anxious children tend to show a higher performance in reading as compared with arithmetic. A group of aggressive children and a group of withdrawing children were given tests of reading and arithmetic and the disparity scores compared. There is a strong tendency for the former to perform badly in arithmetic and Raven's Matrices. These results are discussed in the light of available evidence

### I.—DESIGN OF THE INVESTIGATION.

IN a previous paper (Lunzer, 1960), we reported an investigation designed to compare the adjustment of three groups of children, selected by class teachers as most aggressive, most withdrawing or 'exceptionally well-adjusted.'

Following on the hypothesis advanced by Lynn (1957), it appeared likely that the withdrawing sample would show a marked superiority in reading as compared with their performance in arithmetic. It seemed possible that the aggressive sample would reach a comparatively higher standard in arithmetic, although here the theoretical grounds were less certain, except insofar as statistical considerations alone might tend to create such a picture if the disparity was clear-cut for the withdrawing. In order to test these hypotheses, all children in the three samples aged 8 and over were given the following tests:

- (1) Schonell's Graded-Word Reading Test (R1).
- (2) Schonell's Silent Reading Test, Form B (R4).
- (3) Vernon's Graded Arithmetic-Mathematics Test (here abbreviated to A/M).
- (4) Raven's Progressive Matrices (1938).

The groups covered by the present inquiry consisted of 30 'aggressive' (mean CA, 10 yrs. 4 mths, s.d. 19.08 mths.), 26 'withdrawing' (mean CA, 10 yrs. 5 mths., s.d. 16.3 mths.) and 32 'well-adjusted' children (mean CA, 10 yrs. 5 mths, s.d. 16.9 mths). The method of selection is described more fully in the previous paper. The samples cannot be taken as fully representative (but at the same time the population from which they were drawn was not highly selected).

### II.—RESULTS.

The method of scoring was determined by the fact that we were concerned not with the absolute standards attained by children, but rather with their relative attainments on one test as compared with another. It could not be assumed that the original standardisation samples were comparable for all four tests and hence it did not seem legitimate to compare attainment ages. The procedure finally adopted was to convert all raw scores into sigma scores, using

the subjects of the investigation as an *ad hoc* standardisation sample. Hence, the scores referred to below have no meaning outside the samples on which they were obtained, but they are comparable as between one test and another. The overall mean score on each of the tests is zero and the standard deviation is 1. Table 1 shows the mean scores together with the standard deviations of the three groups taken separately (expressed in terms of the overall distribution).

TABLE 1  
COMPARATIVE ATTAINMENTS ON FOUR TESTS.

	A (N = 30)	W (N = 26)	WA (N = 32)
R1 .....	M - .1370 $\sigma$ 1.0339	- .2504 1.193	.3334 .7858
R4 .....	M - .2670 $\sigma$ 1.1267	- .0746 1.030	+ .3097 .8486
A/M .....	M - .3037 $\sigma$ 1.0158	- .2110 1.134	+ .4456 .7663
Matr. ....	M - .4223 $\sigma$ 1.0997	- .0381 1.0669	+ .4159 .7218

It may be seen that the well-adjusted group show a uniformly higher performance on all four tests, a result which is not unexpected. Moreover, their scores are less widely distributed than those of the two poorly adjusted groups. The latter differences are significant, as is shown by the application of Bartlett's test of homogeneity.

TABLE 2  
HETEROGENEITY OF VARIANCES IN FOUR TESTS.

	B'	d.f.	significance
R1 .....	7.26	2	.05
R4 .....	3.20	2	.10
A/M .....	6.57	2	.05
Matr. ....	7.65	2	.05

An inspection of the distribution of scores obtained by the three groups showed that the lower variance of the well-adjusted sample is due to the comparative rarity of very low scores, rather than to the concentration of scores about the average of the group. In other words, some of the poorly-adjusted children, and none of the 'exceptionally well-adjusted' were very backward. Needless to say, the latter finding may be solely the result of teachers' selection, although this is unlikely.

Some allowance must be made for lack of homogeneity in interpreting analyses of variance (Lindquist, 1953). Results of these are as follows:

TABLE 3

ANALYSIS OF VARIANCE BETWEEN GROUP MEANS.

	F	significance
R1	2.95	N.S. (.05 - 3.13)
R4	2.74	N.S.
A/M	7.40	.01
Matr.	6.03	.01

Differences between groups are not significant for the two tests of reading. For the arithmetic and matrices tests the level of significance is sufficiently high to warrant some confidence in spite of the lack of homogeneity. An inspection of Table 1 shows the greatest difference to lie between the aggressive and the well-adjusted groups, in favour of the latter.

However, these results can best be considered after an examination of difference scores as between tests:

TABLE 4

DISCREPANCIES IN ATTAINMENT ON FOUR TESTS SHOWN BY THREE GROUPS OF CHILDREN.

	A (N=30)	W (N=26)	WA (N=32)
R1—A/M .....	M+ .1667	— .0392	— .1122
	$\sigma$ .7682	.8280	.6289
R4 —A/M .....	M+ .0367	+ .1370	— .1359
	$\sigma$ .6152	.4976	.4763
Matr.—R1 .....	M— .2853	— .2123	— .0825
	$\sigma$ .9434	.7191	.8002
Matr.—R4 .....	M— .1553	+ .0365	— .1062
	$\sigma$ .9662	.7487	.8312
Matr.—A/M .....	M— .1187	+ .1731	— .0297
	$\sigma$ .9373	.6384	.7625

The two results showing the greatest discrepancies as between the three groups were tested by analysis of variance. Neither of these proved significant. For R1—A/M the F ratio is 1.44 and for Matr.—R1 it is 2.88. The latter is not far short of the .05 level (3.13).

### III.—DISCUSSION.

It is interesting to note that the trend of these results is in direct contradiction to the hypothesis advanced earlier. It may be that the comparatively poor performance of aggressive children in arithmetic and in the matrices test is due to their tendency to carelessness. These children have no desire to conform to adult authority, as is shown in their ratings on this trait (Lunzer, 1960). Withdrawing children, on the other hand, are happiest at repetitive and mechanical work, although their standard even in this type of task is below that of the well



adjusted. It is assumed in this interpretation that performance both in arithmetic and in the first two sections of the matrices test is heavily loaded with the effects of perseverance at a relatively low level. It would follow that the relationship found in this study should not be universal. Where the basic skills in reading and arithmetic have not yet been established, neither can be pursued in a mechanical fashion, and one cannot, therefore, generalise with safety as to which of these subjects will present a child with the greater frustration. It is significant in this connection that the mean discrepancy between R1 and A/M for eighteen aggressive children whose reading and arithmetic ages were both above 9.0 was 0.2317 in favour of reading, while there was little to choose between proficiency in reading and arithmetic for the remaining twelve children who had not attained a reasonable standard in both (mean discrepancy 0.0692). It remains possible that if the comparison is made at a much higher level, between mathematics and literary subjects, children of anxious disposition (including withdrawing children) will show specific failure in the former, not as a result of carelessness, but because, at such a level, the work demanded is no longer of a mechanical nature (Skemp, 1958).

In view of the comments made by Reed and Schonfield (1958) on Lynn's interpretation of his results, the present findings are not so much at variance with these as might be supposed. Lynn found that children who made high scores on two 'anxiety' tests were significantly better at reading than arithmetic. For one of these tests no clinical validation is given, but the other is a test found by Himmelweit and Petrie (1951) to discriminate between neurotics and normal children. Since the former sample was taken from children attending a child guidance clinic, without further specification as to the type of maladjustment, there is little reason to believe that the trend found by Lynn should not be due to the presence of a preponderance of 'aggressive' children among the 'anxious' groups.

It will be noticed in Table 3 that 'aggressive' children tended to perform particularly poorly in the matrices test. It might have been supposed that both groups of poorly-adjusted children would tend to be 'under-achievers' while the well-adjusted would be 'over-achievers' or working to capacity. That this was not so underlines the caution made by Vernon (1958) with regard to the use of group non-verbal tests in the diagnosis of retardation. The marked weakness of the 'aggressive' group in the matrices test *may* have been due to carelessness, but this we do not know.

It remains that the question of disparity in attainment is not a simple one.

The following points need to be taken into account :

- (1) The combined reliabilities of two tests must affect results.
- (2) The comparative difficulty of subjects is affected not only by the specific experience of children (teaching skill, absences, etc.), but also by the standard which they have attained in each. Thus, where the mechanics of reading have not been mastered, there is every reason to believe that the learning of reading is at least as complex a task as the learning of arithmetic. At a somewhat more advanced level, both skills may show a high component of mere mechanical repetition but performance in arithmetic is more liable to errors of carelessness. Finally, at the secondary level we may find that skill in mathematics demands a certain elasticity of approach and flexibility of reasoning which are impaired by blockages due to anxiety. But, whatever the level, individual differences will be important, since the experience of success is the all-important factor in the acquisition of skills.

## IV.—REFERENCES.

- <sup>1</sup> HIMMELWEIT, H. T., and PETRIE, A. (1951). The measurement of personality in children. *British Journal of Educational Psychology*, 21, 9-29.
- <sup>2</sup> LINDQUIST, E. F. (1953). *Design of Experiment in Psychology and Education*. Boston: Houghton Mifflin.
- <sup>3</sup> LUNZER, E. A. (1960). Aggressive and withdrawing children in the normal school. I—Patterns of adjustment. *British Journal of Educational Psychology*, 30, 1-10.
- <sup>4</sup> LYNN, R. (1957). Temperamental characteristics related to disparity of attainment in reading and arithmetic. *British Journal of Educational Psychology*, 27, 62-67.
- <sup>5</sup> REED, G., and SCHONFIELD, D. (1958). Disparity of attainment in reading and arithmetic. *British Journal of Educational Psychology*, 28, 271.
- <sup>6</sup> SKEMP, R. (1958). *An Investigation into Difficulties in Learning Mathematics by Children of Good Intelligence*. Manchester University, Ph.D. Thesis.
- <sup>7</sup> VERNON, P. E. (1958). The relation of intelligence to educational backwardness. *Educational Review*, 11, No. 1, 7-15.

(Manuscript received, March 26th, 1959.)

# A NATIONAL SURVEY OF THE ABILITY AND ATTAINMENT OF CHILDREN AT THREE AGE LEVELS

By D. A. PIDGEON

(*National Foundation for Educational Research.*)

**SUMMARY.** During 1955 some 10,000 children attending maintained or assisted schools in England and Wales were given specially designed tests of reading, arithmetic and non-verbal ability. The children tested formed separate samples in each of three age groups. The age groups and the dates of testing were so chosen that the children in the three samples were :

- (i) Commencing their junior school course (7-8 years).
- (ii) Ending their junior school course ( $10\frac{1}{2}$ - $11\frac{1}{2}$  years).
- (iii) In the last year of compulsory education (14-15 years).

A systematic but unbiased method of sampling was employed which was considered to be effectively random.

The aims of the survey were :

- (i) To obtain reliable estimates of attainments as a base-line for future comparisons.
- (ii) To determine the degree of backwardness in each age group.
- (iii) To compare the levels of attainment of boys and girls, and of children attending schools in urban and rural areas and of various types, e.g., mixed, single-sex, all-age, voluntary.

While the results obtained relevant to the first two aims cannot be evaluated until the repeat of the survey, planned for 1960, is carried out, this paper gives some of the more general findings relating to the levels of attainment achieved by the children in the three age groups, and to the differences in performance observed between children in various sub-groups.

## I.—INTRODUCTION.

In 1955, in accordance with its declared policy, the National Foundation for Educational Research carried out a survey of the ability and attainment of representative samples of children in England and Wales. Before the survey took place its aims were stated as follows :

(i) To estimate average levels and ranges of attainment in reading and arithmetic for the three age groups 7+, 10+, and 14+, in order to discover what are the existing standards and thus to provide a basis for future comparisons.

(ii) To determine the amount of backwardness in each age group and to estimate how much of this backwardness is genuine retardation, due to a variety of factors, and how much is attributable to low general ability (as measured by a non-verbal test of intelligence).

(iii) To compare the levels of attainment of boys and girls, and of children attending schools, in urban and rural areas, and of various types, e.g., mixed, single-sex, all-age, and voluntary.

Some results bearing on the second aim were published in 1956 (Pidgeon and Yates, 1956). The important issues raised by this publication and by the subsequent discussion on the relationship between ability and attainment that appeared in the correspondence pages of *Education* (Burt *et al.* 1956) are still



not entirely resolved. It was clear, however, that the data from the 1955 Survey could not be used to obtain measures of 'under-functioning' and hence an operational definition of backwardness had to be adopted. This meant that the second aim of the survey, quoted above, had to be modified and further that the data on backwardness derived from the survey could become fully intelligible only when compared with similar data obtained by repeating the same tests on a subsequent occasion.

Similarly, the results of the survey relevant to the first aim—the establishment of a base-line for future comparisons—also cannot be evaluated until the repeat of the survey which is planned for 1960.

Accordingly, this article is concerned first of all with a brief description of the survey—the sampling procedure, the tests used and the administrative problems, and secondly, with such results as are relevant to the third aim, namely the comparisons of the performances of children in different sub-groups.

## II.—THE SAMPLES OF CHILDREN.

The survey was confined to grant-aided schools maintained or assisted by local education authorities in England and Wales. It was carried out at three age levels and the populations involved at each level were defined as follows:

*7+ Group*: Children in infants, junior and all-age schools who had reached their seventh birthday but not their eighth by August 31st, 1955.

*10+ Group\**: Children in junior and all-age schools who had reached their tenth birthday but not their eleventh by August 31st, 1954.

*14+ Group*: Children in all-age and secondary schools of all types who had reached their fourteenth birthday but not their fifteenth by August 31st, 1955.

After careful consideration, the dates of testing were fixed for October, 1955, for both the 7+ and 14+ age groups, and May, 1955, for the 10+ group. Thus children were tested (a) at the beginning of their junior school course (actual age range 7:2 to 8:1); (b) at the end of their junior school course (actual age range 10:9 to 11:8); and (c) at the end of the period of compulsory schooling (actual age range 14:2 to 15:1).

## III.—THE SAMPLING PROCEDURE.

There is not space here to give full details of the procedure employed. Briefly, the sampling unit was defined as the school and all children in the appropriate age group in each selected school were tested. The samples of schools were picked directly from the lists (kept at the Ministry of Education) of all the schools in England and Wales covering the defined populations. For the 7+ and 10+ age groups, in order to ensure the adequate representation of schools of different types and in urban and rural areas, random sampling was carried out separately in each of these strata. For the 14+ age group, the administrative distinction between urban and rural areas was less important and hence, here, stratification was by sex of school in addition to type. The sampling method was thus systematic but unbiased and was considered to be effectively random.

Tables 1 and 2 give the number of local education authorities involved and the number of schools in each age group. It will be noticed that one English county council education authority refused to participate in the survey. It was not thought, however, that this made any material difference to the representativeness of the results.

\* This is, of course, the age group more usually known as 11+.

TABLE 1  
NUMBER OF L.E.As. TAKING PART IN THE SURVEY.

	Nos. of L.E.As. with schools in			No. of L.E.As. with no schools	Totals
	3 age groups	2 age groups	1 age group		
English C.Cs. ....	15	18	12	4	49*
English C.Bs. ....	1	8	24	46	79
Wales .....	3	3	6	5	17
Total .....	19	29	42	55	145*

\* One English C.C. refused to participate in the Survey.

TABLE 2  
NUMBER OF L.E.As. AND SCHOOLS IN THE THREE AGE GROUPS.

	7+ Group		10+ Group		14+ Group	
	L.E.As.	Schools	L.E.As.	Schools	L.E.As.	Schools
English C.Cs.	32	58	37	62	24	42
English C.Bs.	12	14	16	17	15	17
Wales .....	5	8	10	12	6	8
Total ..	49	80	63	91	45	67

The numbers of children tested were approximately 2,850 at 7+, 3,200 at 10+ and 3,650 at 14+. In the results to be given later, the actual numbers vary, since not all children were present for every test.

#### IV.—THE TESTS GIVEN.

The 7+ group was given a 60-item non-verbal picture test and specially designed tests of reading comprehension and mechanical arithmetic. The picture test was the N.F.E.R. Picture Intelligence Test 1 and its reliability calculated by Kuder-Richardson formula 20 was .93. The arithmetic test consisted of four separate sub-tests of 20 items each, in addition, subtraction, multiplication, and division, administered untimed on different occasions. The Kuder-Richardson reliability of the total test of 80 items was .97. The reading test contained 60 sentence completion items and was the same as that given to children in the older age groups. Unfortunately, although it proved to be highly successful with children of 10+ and 14+, it was found not to give effective discrimination with the younger children and hence had to be omitted from the analyses.

The 10+ group completed four tests; the 60-item reading comprehension test (K-R reliability .96) a 64-item non-verbal test (K-R reliability .94) and two tests of arithmetic. The first of these (Mechanical Arithmetic I, K-R reliability .92) was a 25-item mechanical test ranging from simple addition to multiplication and division of fractions and decimals, given without time limits. The second (Problem Arithmetic II, K-R reliability .98) consisted of 100 items, 40 mechanical and 60 problem, administered with a time limit of 50 minutes. This test was, in fact, similar to the usual type of arithmetic test

used by many local education authorities as part of their 11+ allocation procedures.

The 14+ group completed the 60-item reading comprehension test (K-R reliability, with this age group, .96) and the 25-item untimed mechanical arithmetic test given to the 10+ age group (K-R reliability with this age group .97). In addition, they were given a 96-item non-verbal test (K-R reliability .96) and an 85-item timed problem arithmetic test (Problem Arithmetic III, K-R reliability .97) which included items common to all secondary school curricula.

The tests in all three age groups were given on the mornings only of three consecutive days.

#### V.—ADMINISTRATIVE ARRANGEMENTS.

The administration of a survey of this kind is exceedingly complex and only the barest outline can be given here. In order to ensure uniform conditions for testing, every participating local education authority was asked to appoint a representative who was to be responsible for the local administration. These representatives were met at seventeen regional meetings, scattered over England and Wales, by officers of the Foundation, who briefed them on all details of the survey. In many cases the representatives, often psychologists, carried out the testing themselves. In other cases they in turn briefed the teachers in the schools concerned for this task. Full details of the test administration were given in instruction booklets. The testing was completed by all schools within a two-week period.

It was decided to test on one occasion only, and children who were absent for all tests were ignored. Each test was standardised to a mean of 100 and S.D.15 on the sample formed by all children who took it, although the joint analysis of the results was performed only on the sample of children in each age group who took all tests. The scores of the 150 or so 'partial' absentees were, in fact, examined and found to average some four points of standardised score below the mean; their omission from the main group, however, had a trivial effect on the mean.

#### VI.—RESULTS.

As has already been mentioned, there is little that has any real meaning to be reported at this stage on the question of backwardness. The proportions of children scoring below certain arbitrarily defined points can be calculated, but these have no absolute value and only assume importance when it can be shown that after a passage of time they have changed. However, two of the tests were given to more than one age group and the overlap in performance at different ages is of passing interest.

(i) *Over-lap in performance between age groups.* The same reading test was given to children in all three age groups. As was indicated earlier, this test proved too difficult for the seven-year-olds to allow of any detailed analyses to be made with this age group, but the mean score obtained was found to be just over 7 points. The proportion of children in the 10+ age group scoring 7 or below was found to be 4.0 per cent., while the corresponding proportion in the 14+ age group was 1.4 per cent. It was also found that 17.0 per cent. of children aged 14+ scored less than the 10+ mean, while 20.3 per cent. of the 10+ age group scored more than the 14+ mean.

In arithmetic, the only comparison of this kind that can be made is between the 10+ and 14+ age groups. On the mechanical arithmetic test given to both age groups 25.7 per cent. of the older children scored less than the 10+ mean,



while no fewer than 42.0 per cent. of the 10+ group scored above the 14+ mean. The higher proportions for this test can be accounted for by the fact that it was easier for both age groups; the distributions being negatively skewed and the means relatively closer together. The test was, in fact, included in the survey primarily in order that a study could be made of the responses to individual items. It is intended to report on the results of the analysis of this test on a later occasion.

(ii) *Over-lap in performance between modern and grammar school pupils.* One further comparison of interest is that between pupils in selective and unselective secondary schools. In the reading test at 14+, approximately 4 per cent. of the children in modern and all-age schools scored above the mean for grammar school children. This figure can be compared with that of 5 per cent. quoted in the Ministry's report on reading (Ministry of Education 1957). When this same test was given to the 10+ group, the children were, of course, still in primary schools. Subsequently, however, information was obtained on the type of secondary school to which they were transferred and the extent of the overlap at this age calculated. At 10+, also, approximately 4 per cent. of children who were subsequently transferred to modern and all-age schools scored above the mean for children who later went to grammar schools. In both cases, of course, the figure quoted is only one way of reporting the lack of perfect correlation between the reading test and whatever procedure was used to allocate the children to the selective and unselective schools. It is interesting to note that the figure is unchanged between 10+ and 14+.

Similar calculations were made for the mechanical arithmetic Test 1. Here, however, the extent of the overlap, that is, the percentage of modern and all-age children scoring above the grammar school mean, was 6 per cent. at 10+ and 22 per cent. at 14+. The higher figures presumably again reflect the limited range of content of the test.

(iii) *Sex Differences.* Table 3 gives the mean standardised scores for boys and girls separately on the tests given to all three age groups. The final column gives the difference of the means for the two sexes with standard errors.

TABLE 3

MEAN STANDARDISED SCORES FOR BOYS AND GIRLS SEPARATELY ON TESTS GIVEN TO ALL THREE AGE GROUPS.

Age Group and Test	Boys	Girls	Boys—Girls with S.Es.
7—			
Non-Verbal .....	100.09	99.71	0.38 (0.67)
Mechanical Arithmetic	99.14	100.91	-1.77 (0.66)
10—			
Non-Verbal .....	99.97	101.02	-1.05 (0.56)
Reading .....	99.71	100.39	-0.68 (0.57)
Mechanical Arithmetic I	98.73	102.03	-3.30 (0.63)
Problem Arithmetic II	100.00	100.81	-0.81 (0.60)
14—			
Non-Verbal .....	100.25	101.16	-0.91 (0.71)
Reading .....	100.15	101.12	-0.97 (0.80)
Mechanical Arithmetic I	99.97	100.29	-0.32 (0.63)
Problem Arithmetic III	102.73	97.88	4.85 (0.76)

As will be seen, the differences for the non-verbal tests at all three age groups, and those for the reading tests at 10+ and 14+, do not achieve statistical significance. A brief word of comment can be made about the reading test. Although not significant, the higher mean scores achieved by girls on this test are consistent with the results obtained when this same test has been used in other researches. It will be recalled, however, that the Ministry of Education's Reading Surveys (Ministry of Education, 1957) which used a similar type of reading comprehension test, found a sex difference in favour of boys at both 11 and 15. It is clear that the explanation of these sex differences lies in the content of the tests and not in any real superiority of one sex over the other.

The most interesting results from Table 3 are those for the arithmetic tests. In mechanical arithmetic it will be seen that girls obtain a significantly higher mean score at 7+, that is at the beginning of the junior school course. At the end of the course the difference in favour of girls is even greater. But, after 3 years in secondary schools the difference has all but disappeared. In tests involving problem arithmetic, girls at 10+ obtain a higher mean score than boys, although the difference is not significant. By 14+ however, boys' superiority is nearly 5 points of standardised score. These results indicate that there is distinct falling off in the performance of girls in this subject between the ages of 11 and 14.

It is too often assumed, especially by secondary school teachers, that boys are inherently capable of better performance in mathematics than girls, although no valid evidence for this assumption exists. Certainly, the fact that at 14 boys are superior to girls cannot be used as evidence to support this contention, since at the age of 11 and even as young as 7, the opposite is the case.

From the item analyses carried out on the mechanical arithmetic tests, the average percentages of boys and girls in both the 10+ and 14+ groups answering correctly the 25 items of the mechanical arithmetic Test 1 were calculated, and are given in Table 4.

TABLE 4

AVERAGE PERCENTAGES OF BOYS AND GIRLS IN DIFFERENT TYPES OF SECONDARY SCHOOL AT 10+ AND 14+ ANSWERING ITEMS ON MECHANICAL ARITHMETIC TEST 1 CORRECTLY.

	Secondary Modern and All-age		Selective Schools		All types	
	10+	14+	10+	14+	10+	14+
	%	%	%	%	%	%
Boys .....	50.2	70.0	85.5	86.8	58.0	73.3
Girls .....	55.9	68.2	87.5	88.2	63.6	73.0
Difference: Boys—Girls ..	-5.7	+1.8	-2.0	-1.4	-5.6	+0.3

This table presents these percentages separately for children attending selective and non-selective secondary schools. The data at 10+ were, of course, obtained when the children were still in primary schools, but were analysed according to the type of secondary school the children subsequently attended. It is interesting to note that the superiority of grammar school girls

is more or less maintained between 11 and 14, and it is in the Modern and All-age schools that the girls appear to have done relatively badly.

Clearly, a further investigation into the teaching of this subject to girls in Modern schools would be interesting, or possibly for an explanation it is necessary to examine the way the subject is taught in primary schools (Dienes 1959).

(iv) *Urban-Rural Differences.* With very few exceptions, large-scale surveys involving a comparison of children's test performance in urban and rural areas have demonstrated the superiority of urban children. (Terman and Merrill 1937, Cross and Revell 1956, Bristol Institute of Education 1952, Morris 1959, Ministry of Education 1957). In general, the data from this survey support these previous findings. The mean standardised scores on the tests given to the 7+ and 10+ age groups for urban and rural children are given in Table 5.

TABLE 5

MEAN STANDARDISED TEST SCORES FOR CHILDREN IN URBAN AND RURAL AREAS IN THE 7+ AND 10+ AGE GROUPS.

Age Group and Test	Urban	Rural	Urban-Rural with S.Es.
7+			
Non-Verbal .....	100.36	97.12	3.24 (1.23)
Mechanical Arithmetic	99.79	101.27	-1.48 (1.61)
10+			
Non-Verbal .....	100.91	98.06	2.85 (1.54)
Reading .....	100.61	96.85	3.76 (1.55)
Mechanical Arithmetic I	100.84	97.47	3.37 (1.45)
Problem Arithmetic II	100.90	97.56	3.34 (1.44)

The 7+ mechanical arithmetic test results are an interesting exception to the general trend, but the difference is not statistically significant. In the attainment tests at 10+, the rural children come a very poor second.

However, as Barr (1959) has pointed out, the main influence at work in producing results such as these is not the nature of the rural or urban environment itself, but rather "the structure of the socio-economic group from which any school draws its pupils."

The explanation of the differences in Table 5 then, may well be that there is a high positive correlation between test scores and home environment (Fraser, 1959) and there are considerably fewer 'better' homes in rural areas.

(v) *Designation of School.* No stratification by designation of school was made in the sampling procedure, but after the samples of schools were known, local authorities were asked to state whether it was a county or voluntary school. Using this information, the mean standardised scores for children attending the different schools were determined. Since the numbers in the different types of voluntary schools were fairly small, they have been grouped together for comparison with County schools. In fact, only trivial differences were revealed between these types. The results are given in Table 6.



TABLE 6

MEAN STANDARDISED TEST SCORES FOR CHILDREN ATTENDING COUNTY AND VOLUNTARY SCHOOLS IN THE 7+ AND 10+ AGE GROUPS.

Age Group and Test	County Schools	Voluntary Schools	County-Voluntary with S.Es.
7+			
Non-Verbal .....	101.08	96.33	4.75 (1.36)
Mechanical Arithmetic	100.38	98.84	1.54 (1.67)
10+			
Non-Verbal .....	101.72	96.77	4.95 (1.31)
Reading .....	100.91	97.41	3.50 (1.15)
Mechanical Arithmetic I	101.74	96.08	5.66 (1.50)
Problem Arithmetic II	101.87	95.94	5.93 (1.34)

In order to account for the highly significant differences that were found, the hypothesis can be put forward that voluntary schools are more likely to be in poorer social areas.

Unfortunately, no data on the socio-economic status of the catchment areas for the schools concerned were collected in the National Survey.

(vi) *School Type Differences.* The results of the survey relevant to the differences in performance of 7+ and 10+ children attending junior with infants and junior without infants schools have been published elsewhere (Pidgeon 1959). They can conveniently be summarized here by saying that although the mean standardised scores in nearly all cases favoured the latter type of organisation, when further analyses were carried out in an attempt to eliminate the effects of the social grouping of the catchment areas for each school, the differences disappeared, or, in the case of reading at 10+, were reversed.

With the 14+ age group differences in performance between children attending modern and grammar schools are, of course, expected, but of greater interest are the differences between children from modern and all-age schools. The mean standardised test scores for children from different types of secondary school are given in Table 7.

TABLE 7

MEAN STANDARDISED SCORES FOR CHILDREN OF 14+ ATTENDING DIFFERENT TYPES OF SECONDARY SCHOOL.

Test	Type of School		
	All-age	Modern	Grammar
Non-verbal .....	94.82 (1.11)	97.33 (0.94)	115.04 (0.97)
Reading .....	92.10 (0.61)	97.17 (1.13)	117.68 (0.81)
Mechanical Arithmetic I ..	97.76 (1.28)	97.54 (0.97)	109.79 (0.84)
Problem Arithmetic III ....	93.96 (1.10)	96.20 (0.99)	117.57 (1.36)

Many theories can be advanced to explain the differences in results between all-age and modern schools. Again, one obvious one is the fact that modern schools, being a relatively newer type of school, are more likely to be found in better areas; greater numbers of all-age schools are still to be found in poorer rural districts. Clearly, further evidence must be obtained if this hypothesis is to be tested.

(vii) *Sex of School.* The last comparison to be made from these survey data is that of the differences in performance of children in the 14+ age group attending single-sex or mixed schools. In this age group the stratification in the sampling, apart from type of school, was on this variable. The mean standardised scores for children attending single sex or mixed schools, whether they were grammar, modern or any other type, are given in Table 8. As will be seen, the differences for all tests except the mechanical arithmetic are significant.

TABLE 8

MEAN STANDARDISED TEST SCORES FOR CHILDREN IN THE 14+ AGE GROUP ATTENDING SINGLE-SEX OR MIXED SCHOOLS.

Test	Single-Sex Schools	Mixed Schools	Single-Sex— Mixed with S.Es.
Non-Verbal . . . . .	102.95	98.77	4.18 (1.35)
Reading . . . . .	103.10	98.51	4.59 (1.53)
Mechanical Arithmetic I . .	100.90	99.48	1.42 (1.22)
Problem Arithmetic III . . .	101.95	98.65	3.30 (1.48)

One possible explanation of the differences may be that where the choice is offered parents still prefer single-sex schools, and since many long-established grammar schools are of this kind, in any 11+ allocation procedure, they fill up first with the brighter pupils. To test this, the performances of children in the two kinds of school were considered separately in the grammar and modern strata. The results, however, showed no clear pattern and certainly gave no indication that this was the correct or only explanation. It is also possible, of course, that an explanation of the differences reported in Table 8 lies in the fact that single sex schools tend to be in better areas.

#### VII.—CONCLUSION.

The aim of this paper was to present some of the results obtained from administering tests to fully representative samples of children from the whole of England and Wales. For the most part, the results reported here have been confined to the comparisons of certain sub-groups, and in doing this far more problems have been raised than solved. The survey which provided the data was not designed to provide evidence to explain any differences found; its intended purpose was merely to discover whether differences existed or not.

In presenting the results the attempt has been made in some cases to offer certain explanations of discovered differences in performance between sub-groups of children, but it is not suggested that these explanations are the correct or only ones possible. In the case of children from urban and rural areas, it seems probable that the differences in performance found can, to a very large extent, if not completely, be accounted for by considering social and cultural factors associated with the home backgrounds of the children concerned. It may well

be true that similar factors affect many of the other differences reported in this paper, but certainly more exact evidence is required before a claim of this kind can be substantiated. In 1960, the National Foundation intends to repeat its survey and on that occasion it is hoped to obtain additional information which may help to answer the many questions still outstanding.

#### VIII.—REFERENCES.

- BARR, F. (1959). Urban and rural differences in ability and attainment. *Educational Research*, Vol. 1, No. 2.
- BRISTOL—INSTITUTE OF EDUCATION (1952). *Studies in Selection Techniques for Admission to Grammar Schools*. Publication No. 3. University of London Press.
- BURT, C., *et al.* (1956-7). Correspondence in *Education*, Dec., 1956, to March, 1957.
- CROSS, G. R., and REVELL, K. J. (1956). Note on grammar school selection in a rural area. *National Foundation for Educational Research Bulletin*, No. 7.
- DIENES, Z. P. (1959). On the growth of mathematical concepts in children through experience. *Educational Research*, Vol. 2, No. 1.
- FRASER, E. (1959). *Home Environment and the School*. University of London Press.
- MINISTRY OF EDUCATION (1957). *Standards of Reading*. H.M. Stationery Office.
- MORRIS, J. M. (1959). *Reading in the Primary School*. London: Newnes.
- PIDGEON, D. A. (1959). School type differences in ability and attainment. *Educational Research*, Vol. 1, No. 3.
- PIDGEON, D. A., and YATES, A. (1956). The relationship between ability and attainment. An examination of current theory. *National Foundation for Educational Research, Bulletin No. 8*.
- TERMAN, L. M., and MERRILL, M. A. (1937). *Measuring Intelligence*. London: Harrap.

(Manuscript received, September 19th, 1959.)



# THE MEASUREMENT OF EFFICIENCY OF SELECTION BY PROCEDURES OF VARYING VALIDITIES

By FRANK SANDON

(Sometime Selection Officer, Birmingham Education Committee.)

**SUMMARY.** Tables are given to show the numbers and percentages 'correctly' chosen at various intensities of selection by selection procedures of various validities. The tables are derived on the assumption of bivariate normal distribution, i.e., on the assumption that the measures of both the selection procedure and of the criterion are normally distributed, that the regressions are linear, and that the various arrays of each measure are homoscedastic.

## I.—INTRODUCTION.

In *Statistics in Psychology and Education*, Garrett<sup>1</sup> gives a table (Table 23, quoting from Taylor and Russell<sup>6</sup>), showing the proportions considered satisfactory in achievement in the criterion range from .05 to .95, where the correlations between test scores and criterion may range from .00 to 1.00. The table does not appear so convenient as alternatives that can be derived from the Tables for the Volumes of the Normal Bivariate Surface<sup>2</sup> prepared by co-workers of Karl Pearson. These latter tables give, for various values of  $r$  (the correlation coefficient), the volume of the sector of the bivariate normal surface cut off at different values of  $h$  and of  $k$ , the two independent variables. From these, Tables 1 and 2 have been derived.

TABLE 1

NUMBERS 'CORRECTLY' CHOSEN BY TESTS CORRELATING  $r$  WITH CRITERION AND INTENDED TO SELECT THE TOP  $p$  PER CENT. FROM A POPULATION OF 10,000.

$r$	$p$							
	5	10	15	20	25	30	35	40
1.00	500	1000	1500	2000	2500	3000	3500	4000
.95	371	779	1206	1647	2099	2560	3031	3512
.90	319	689	1085	1500	1931	2376	2835	3306
.85	280	620	992	1387	1802	2234	2682	3147
.80	248	563	913	1291	1692	2113	2552	3010
.75	220	513	845	1207	1594	2005	2436	2889
.70	196	468	783	1130	1505	1906	2330	2778
.65	175	428	726	1059	1423	1815	2231	2671
.60	156	391	673	994	1346	1729	2138	2576
.55	138	356	624	932	1273	1647	2050	2483
.50	122	324	578	873	1204	1569	1965	2393
.45	108	295	534	817	1138	1494	1883	2332
.40	95	267	493	764	1074	1421	1804	2223
.35	83	241	454	712	1012	1351	1727	2141
.30	72	217	416	663	953	1283	1652	2061
.25	62	194	381	616	895	1216	1578	1982
.20	53	172	347	570	838	1151	1506	1905
.15	45	152	314	526	783	1087	1435	1829
.10	37	134	284	483	730	1024	1365	1718
.05	31	117	254	442	678	963	1295	1678
.00	25	100	225	400	625	900	1225	1600

TABLE 2

PERCENTAGES 'CORRECTLY' CHOSEN OF ALL CHOSEN BY TESTS CORRELATING  $r$  WITH CRITERION and INTENDED TO SELECT THE TOP  $p$  PER CENT. OF A POPULATION.

$r$	$p$							
	5	10	15	20	25	30	35	40
1.00	100	100	100	100	100	100	100	100
.95	74	78	80	82	84	85	87	88
.90	64	69	72	75	77	79	81	83
.85	56	62	66	69	72	74	77	79
.80	50	56	61	65	68	70	73	75
.75	44	51	56	60	64	67	70	72
.70	39	47	52	57	60	64	67	69
.65	35	43	48	53	57	60	64	67
.60	31	39	45	50	54	58	61	64
.55	28	36	42	47	51	55	59	62
.50	24	32	39	44	48	52	56	60
.45	22	29	36	41	46	50	54	58
.40	19	27	33	38	43	47	52	56
.35	17	24	30	36	40	45	49	54
.30	14	22	28	33	38	43	47	52
.25	12	19	25	31	36	41	45	50
.20	11	17	23	28	34	38	43	48
.15	8.9	15	21	26	31	36	41	46
.10	7.5	13	19	24	29	34	39	44
.05	6.2	12	17	22	27	32	37	42
.00	5.0	10	15	20	25	30	35	40

## II.—USE OF THE TABLES.

For the educational psychologist, the chief function of the tables is to indicate how selection actually works in, for example, selecting the brighter pupils for the more exacting of various courses, when the proportion of the whole population to be chosen for such courses is specified and the correlation between the selecting procedure and the criterion is known. Thus, if we wish to select the best 15 per cent. of a complete age group of 10,000 children by tests for which the coefficient of correlation (over the whole age group) with later success is .80, then we shall select 1,500. From Table 1 we see that, of these, there are 913 who are 'correctly' chosen, and 587 who are 'incorrectly' chosen in place of a different group of 587 who are 'incorrectly' omitted. Similarly, Table 2 shows that 61 per cent. (more precisely 60.9 per cent). are 'correctly' chosen. Table 3 summarizes the position.

TABLE 3

By Criterion	Selection Procedure		
	Omits	Selects	Total
Not Suitable .....	7913	587	8500
Suitable .....	587	913	1500
Total .....	8500	1500	10000

On the whole, if such a result were one in an '11+' examination,<sup>3</sup> then the contributory primary schools would, as pointed out in 1936,<sup>3</sup> be fairly satisfied. Out of every twenty of their pupils the sixteen more unsuitable ones are not chosen and two of the more suitable ones are. There are two other cases: here the examination seems to have chosen the one who is the poorer and omitted the better (of course, at the border line there is really little to choose between these latter two). But the receiving grammar schools would find that out of every five of their new pupils three only are really fitted for the course.

### III.—SOME CONSEQUENCES OF THE TABLES.

These features of the tables often seem to cause surprise among those who are not fully familiar with the nature of the bivariate normal surface,<sup>4</sup> and to cause grave dissatisfaction with the selection procedure among statistically illiterate politicians and members of education committees. Actual experience, however, shows that we always have imperfect correlation between the selecting procedure and the criterion. This is true whatever the selecting procedure, whether, e.g., by tests, by interviews, by impressions, or by school records. If we classify and reclassify, term by term on school reports, as we do in the procedure that has been called 'progressive differentiation,' we still have this imperfect correlation. Even if in this last case we could get a true classification at the beginning of the school term, individual differences in the rate of the attainment of intellectual or academic maturity would mean that at the end of the term the classification would no longer be true—there would be imperfect correlation between the two rankings. In consequence, in any practical case of selection, we shall inevitably choose some 'misfits' and reject some of the 'fit,' and, in the majority of cases, these 'errors' can be disturbingly numerous.

The question of 'sets' as a matter of school organisation illustrates a corollary of these conclusions. In certain cases, as for subjects that are more or less specific to one sex or type of course (such as P.E., handicraft, or Latin *versus* German), statistical considerations hardly enter. If there is a subject that is highly specific in its factors, i.e., one that correlates only slightly with all the other subjects of a school course, then, if it is measured term by term with fair reliability (as, perhaps, may be the case for music), 'setting' may be justified. But, in other cases, such as those where 'setting' has frequently been recommended by official opinion (e.g., for Mathematics, or French) where the abilities are not highly specific (i.e., where the correlations of the marks for the particular subject with the marks for all the other subjects are high) the possibility of 'misfits' in classifying, by form or by 'set,' is determined by the term-by-term correlation of the marks of such subjects or group of subjects. Experience in grammar schools gives a figure of about .8 for these correlations (slightly higher, of course, for all the subjects combined, slightly less for a single one). Hence, for a three-form entry school (e.g., of about 100 pupils per year), if we try to get a form or 'set' of, say, thirty-five, there will be only twenty-five pupils 'correctly' placed in the top or in the bottom form or 'set.' With ten 'misfits' per form or 'set' each term it hardly seems justifiable to have the two classifications—for form work and for 'set' for special subjects—complicating the time table.

It should also be noticed that we can obtain a larger proportion of the 'fit' in any selection procedure only by reducing our standard of 'fitness' and, in consequence, by increasing the actual number of the 'unfit.' Thus, if we are dissatisfied with getting only 61 per cent. 'fit' in 1,500 children chosen out of population of 10,000 when we use a test for which the correlation is .80, we can choose 75 per cent. 'fit' by looking for 4,000 of the 10,000. But our standard of



'fitness' has gone down and we are actually admitting 4,000—3,010, or 990, who are 'unfit' even at this low level, in place of our former 1,500—913, or 587, 'unfit' at the more stringent level.

#### IV.—SOME OTHER APPLICATIONS OF THE TABLES.

The tables, while appropriate for practical problems in education as indicated above, are also suitable for use in other kinds of problem. Thus suppose that we are choosing, for any purpose, at a level that will select 15 per cent. of all twin children. Then the twin pairs may be split by the selection procedure. If the selection procedure correlation between the two individuals of each pair of fraternal like-sex ('two-egg') twins is .60,<sup>5</sup> we have the state of affairs shown in Table 4. There are only 673 twins that are selected, together

TABLE 4

Twins Y	Twins X		
	Not Selected	Selected	Total
Not Selected .....	7673	827	8500
Selected .....	827	673	1500
Total .....	8500	1500	10000

with their co-twins, out of every 1,500 twins selected. For the identical ('one-egg') twins, the correlation is higher. If we assume that the selection procedure frequency distribution of all the various kinds of twins is the same (an assumption that is not exactly true when we select by an examination<sup>5</sup>), then Table 3 can be read to show that out of every 1,500 identical twins that are selected, there will be 913 that are selected with their co-twins. In other words, the receiving schools will find that there are proportionately more identical twins on their roll than fraternal twins than would be found in the complete age group. This feature works at the other end as well. There will be a larger proportion of identical twins to be found in the E.S.N. schools than in the ordinary schools. A further feature would be if we consider, instead of selection by an examination, selection by death. This will split the fraternal twin pairs more than it will split the identical twin pairs, so that the number of identical twin pairs will gradually increase proportionately through life when compared with the number of fraternal twin pairs.

The tables can similarly be used for other cases of selection.

#### V.—REFERENCES.

- <sup>1</sup> GARRETT, HENRY E. (1958). *Statistics in Psychology and Education*. 5th Edition. London: Longmans, Green and Co.
- <sup>2</sup> PEARSON, KARL (Editor) (1931). *Tables for Statisticians and Biometricians, Part II*. London: University College.
- <sup>3</sup> SANDON, FRANK (1936). Selection by a nearly perfect examination. *Annals of Eugenics*, 7 (1) 65-85.
- <sup>4</sup> SANDON, FRANK (1955). A note about a paper on some aspects of selection for grammar schools. *Educational Review*, 7 (2), 150-152.
- <sup>5</sup> SANDON, FRANK (1957). The relative numbers and abilities of some ten-year-old twins. *J. of the Royal Statistical Society, Series A, General*, 120 (4), 440-450. See also SANDON, FRANK (1959). Twins in the school population. *Brit. J. Statist. Psychol.*, 12, 133-138.
- <sup>6</sup> TAYLOR, H. C., and RUSSELL, J. T. (1939). The relationships of validity coefficients to the practical effectiveness of tests in selection. *J. of Applied Psychology*, 23, 565-578.

(Manuscript received, August 12th, 1959.)

# THE VALIDITY OF TESTS OF SPATIAL ABILITY AS PREDICTORS OF SUCCESS ON TECHNICAL COURSES

By I. MACFARLANE SMITH

(*King's College, University of Durham.*)

**SUMMARY.** The object of this investigation was to assess the validity of two tests of spatial ability against criteria of success on technical courses in a secondary technical school.

Scores on the usual selection tests as well as on a test of spatial ability were correlated with marks obtained in internal and external examinations in a number of technical subjects taken (i) three years, and (ii) five years after the original selection examination.

## I.—INTRODUCTION.

THERE is already extensive evidence that scores on spatial tests administered during the secondary school stage are predictive of proficiency in technical occupations, e.g., Holliday<sup>6</sup> and Shuttleworth.<sup>11</sup> The two tests to be discussed in the present article, (i) Spatial Test I (N.F.E.R.) and (ii) Moray House Space Test I, were constructed in order to measure the spatial ability of pupils in the age ranges 11 : 0 to 13 : 11 and 10 : 0 to 11 : 11, respectively. As both of these tests may be regarded as pioneering 'efforts,' it may be of some interest to give a brief account of their history and origins.

A full account of the first of these tests (Spatial Test I) has already been published in an article which included detailed information regarding its content and reliability together with some preliminary evidence of its validity.<sup>18</sup> The test was originally constructed in its main essentials early in 1934<sup>14</sup> and was printed in 1950 on behalf of N.F.E.R. as Spatial Test I. It was fully standardised over the age range 11 : 0 to 13 : 11 on three complete age-groups totalling some 8,000 pupils, (Norms are now available for slightly younger children.) The test has been used by a number of local authorities as well as by numerous research workers (e.g., 21, 23). In the preceding years, research on somewhat similar lines had been carried out by a number of other workers such as El Koussy<sup>7</sup>, Drew,<sup>4</sup> Peel,<sup>10</sup> Bradford,<sup>2</sup> Dempster,<sup>3</sup> Emmett,<sup>5</sup> Watts and Slater,<sup>20</sup> and Watts.<sup>21</sup>

Work on Moray House Space Test I was initiated in 1944 by the late Professor Godfrey Thomson, assisted by J. H. Gray and J. Y. Erskine, and was continued by W. G. Emmett in conjunction with J. T. Bain and L. F. Mills. A factor-analysis of test-scores was begun by Professor Godfrey Thomson, carried further by Swanson, and completed by W. G. Emmett, using D. N. Lawley's Method of Maximum Likelihood. Emmett found four significant factors and tentatively interpreted the first three of these as *g*, *v* and *k*.

Meantime, the writer had included Moray House Space Test I in the battery of tests which he administered to some 200 eleven-year-old boys in June and July, 1947. Like W. G. Emmett, he first applied Thurstone's Centroid Method to the resulting correlations and then D. N. Lawley's Method of Maximum Likelihood. Using Lawley's test of significance he found only two significant factors, but he interpreted the second factor as a bipolar factor, the positive values indicating verbal loadings and the negative values spatial loadings. According to this interpretation, the draft of Spatial Test I had a very substantial spatial loading, but both sections of Moray House Space I had relatively low spatial loadings. (c.f. <sup>18</sup>) If the same interpretation is given to the second factor

extracted from the Moray House correlation matrix by Lawley's method,<sup>5</sup> the same conclusion would follow regarding M.H. Space Test I, i.e., both sections would appear to have relatively low spatial loadings. If, however, following Emmett, the third factor is interpreted as the spatial factor, both sections of the test and particularly the three-dimensional section would appear to have very substantial spatial loadings.

In view of the conflicting interpretations which have been given to the results of these factor-analyses, it was thought that it would be of considerable interest to compare the predictive values of the two spatial tests, when assessed under approximately the same conditions. The data were made available through the kindness of Mr. T. H. C. Walker, now Headmaster of Tynemouth Technical School, and Mr. Stanley Hirst, Director of Education for Middlesbrough. The criteria, however, do not relate to the same samples of pupils. The 'follow-up' for Spatial Test I was carried out with a 13-year-old group after a period of three years, that for M.H. Space Test I was mainly with an 11-year-old group after a period of five years.

So far as the writer is aware, these periods are longer than those for which results have been previously obtained. According to the manual for M.H. Space Test I, assessments for seventy-one grammar school boys in geometry, science and art by the head teacher at the end of the first year were found to have higher correlations with the scores in M.H. Space Test I than with the verbal intelligence, English and arithmetic scores in the admission examination. Also in a secondary modern school, the correlations of M.H. Space Test I with assessments of forty-four boys after one year's work in science, technical drawing, woodwork and metalwork were found to be very satisfactory. In the present enquiry, the results relating to M.H. Space Test I show a somewhat similar trend.

## II.—TESTS USED IN THE SELECTION PROCEDURE IN MIDDLESBROUGH IN 1950 AND 1951

For the present investigation, a 'follow-up' was carried out of pupils admitted to the Technical School in Middlesbrough in 1950 and 1951. The sample of pupils studied consisted of three groups :

(i) A group of sixty 13-year-old pupils who had been tested for re-allocation at 13 in 1951. These had taken the following tests :

M.H. Intelligence (Adv. 9).

M.H. English 18.

M.H. Arithmetic 18.

N.F.E.R. Spatial Test I (Macfarlane Smith).

(ii) A group of ninety 13-year-old pupils who had been tested for re-allocation at 13 in 1950. These had taken the following tests :

M.H. Intelligence 29.

M.H. English 12.

M.H. Arithmetic 12.

M.H. Space Test I (Bain and colleagues).

(iii) A group of sixty 11-year-old pupils who had been admitted to the Technical School after taking the following tests in 1951 :

M.H. Verbal Reasoning 46.

M.H. English 22b.

M.H. Arithmetic 22.

M.H. Space Test I (Bain and colleagues).



The third battery of tests had been given to the whole of the 11 plus age-group. Together with teachers' estimates and school recommendation lists, this had formed the basis of the selection procedure for all the selective schools in Middlesbrough. Pupils had been selected in the first place either for the Technical School or for the Hugh Bell Selective School on the basis of scores in the verbal intelligence, English and arithmetic tests, together with teachers' estimates and school recommendations.

Differences between the spatial test score and the verbal intelligence test score, considered in conjunction with the attainment test scores, were regarded as giving an indication of the suitability of a boy for one or other of these two schools. It was considered that a good spatial test score associated with a good arithmetic score was likely to indicate a greater measure of suitability for the Technical School, while on the other hand, a relatively good verbal intelligence score with a good English score was thought to indicate greater suitability for the Hugh Bell School.

### III.—CORRELATIONS WITH EXAMINATION RESULTS.

Correlations with subsequent examination results in various technical subjects are shown in Tables 1 and 3. Since the tests being evaluated had been taken into account in the selection procedure, neither of them can be expected to reveal its full predictive validity. Thus, the spatial tests would have shown higher correlations with the criteria if the low scoring pupils had been admitted to the Technical School, and similar considerations apply, of course, to the other

TABLE 1  
VALIDITY OF SPATIAL TEST 1 (N.F.E.R.).

Follow-up after three years of pupils who took a battery of selection tests in 1951, when they were 13 years of age.

Correlations are shown between scores in individual tests and criteria of success in technical examinations taken three years later.

Subject	N	Spatial Test 1 (N.F.E.R.)	M.H. Intelligence Test Adv. 9 (Verbal)	M.H. English Test 18	M.H. Arithmetic Test 18
Engineering Drawing (Internal exam., 1954)	52	+ .622	+ .027	— .003	— .106
Metalwork (Theory and practical. Internal exam., 1954) . . . . .	52	+ .346	— .048	+ .150	— .299
Woodwork (Theory and practical. Internal exam., 1954) . . . . .	52	+ .474	— .178	— .014	— .151
Metalwork (G.C.E. J.M. Board, 1954) . . . . .	19	+ .335	— .313	+ .146	— .004
Woodwork (G.C.E. J.M. Board, 1954) . . . . .	24	+ .234	— .099	+ .343	— .223
Geometrical Drawing (G.C.E., J.M. Board, 1954) . . . . .	19	+ .037	+ .559	— .105	+ .693

tests used in the admission examination. Since results were available only for the selected sample of pupils, the correlations are considerably lower than they would have been, if data had been available for an unselected sample. Corrections have not been applied, however, to allow for the attenuation which has resulted from the reduction in the range of scores, and it cannot be assumed that all tests have been equally affected. Regression coefficients and the corresponding multiple correlations with the various criteria are given in Tables 2 and 4.\*

Certain tentative conclusions can be drawn from the results. The correlations of the spatial tests with the criteria are in nearly all cases higher, and in some cases substantially higher, than the correlations of the other tests used in the admission examination. It is, therefore, probable that in selecting pupils for courses of technical education, the inclusion of a spatial test in the selection battery would add to its predictive value and would result in the selection of a more promising sample of pupils. A comparison of the correlations for the two spatial tests suggests that Spatial Test I has, on the whole, higher validities after three years than has M.H. Space Test I after five years, but this conclusion must be regarded as only tentative since the selection procedures at 11+ and 13+ may not have been identical. Regression coefficients for both spatial tests were usually substantial and positive, while more than half of the regression coefficients for the verbal intelligence tests were negative.

TABLE 2  
VALIDITY OF SPATIAL TEST 1 (N.F.E.R.).

Best weightings of test-scores in the selection battery for predicting subsequent success in the examinations. The multiple correlation is given in the last column and the regression coefficients in the preceding four columns.

Subject	N	Spatial Test 1 (NFER)	M.H. Intell. Test Adv. 9 (Verbal)	M.H. English Test 18	M.H. Arithmetic Test 18	Multiple Correlation with Criterion
REGRESSION COEFFICIENTS						
Engineering Drawing (Internal exam., 1954) .. . . .	52	+ .833	- .036	+ .024	- .192	+ .705
Metalwork (Theory and Practical. Internal exam., 1954)	52	- .279	+ .060	- .143	- .283	+ .447
Woodwork (Theory and Practical. Internal exam., 1954)	52	- .603	- .538	- .048	- .122	+ .603
Metalwork (G.C.E. J.M. Board, 1954) .. . . .	19	+ .502	- 1.250	+ .214	+ .362	+ .768
Woodwork (G.C.E. J.M. Board 1954) .. . . .	24	+ .140	+ .104	+ .414	- .102	+ .433
Geometrical Drawing (G.C.E. J.M. Board, 1954) .. . . .	19	+ .242	+ .565	+ .180	+ .641	+ .866

Mean multiple correlation = + .637.

\* The correlations and multiple regression coefficients shown in Tables 1, 2, 3 and 4 were calculated on a Pegasus Computer using the Multiple Regression Analysis Programme Mark IV, prepared by S. A. Robinson and R. J. Taylor (British Iron and Steel Research Association, List No. 101, 1959).

TABLE 3

## VALIDITY OF MORAY HOUSE SPACE TEST 1.

(a) Follow-up after five years of pupils who took a battery of selection tests in 1951 when they were 11 years of age. Correlations are shown between scores in individual tests and criteria of success in technical examinations taken five years later.

Subject	N	M.H. Space Test 1	M.H. Intelligence Test 46 (Verbal)	M.H. English Test 22b	M.H. Arithmetic Test 22
Engineering Drawing (Internal exam., 1956)	47	+ .408	— .071	— .263	+ .030
Engineering Drawing (G.C.E. Assoc. Exam. Board, 1956) . . . . .	27	+ .298	— .154	+ .047	+ .063
Metalwork (G.C.E. Assoc. Exam. Board, 1956) . .	21	+ .187	+ .064	— .126	+ .071
Woodwork (G.C.E. Assoc. Exam. Board, 1956) . .	23	— .072	— .168	+ .042	+ .101
Metalwork (Theory and practical. Internal exam., 1956) . . . . .	39	+ .283	— .033	— .259	— .087

(b) Follow-up after three years of pupils who took a battery of selection tests in 1950 when they were 13 years of age.

Subject	N	M.H. Space Test 1	M.H. Intelligence Test 46 (Verbal)	M.H. English Test 12	M.H. Arithmetic Test 12
Woodwork (Theory and practical. Internal exam., 1953) . . . . .	86	+ .325	— .109	— .212	— .003
Woodwork (Practical only. Internal exam., 1953) . . . . .	88	+ .293	— .029	— .133	— .020



TABLE 4

## VALIDITY OF MORAY HOUSE SPACE TEST 1.

Best weightings of test-scores in the selection battery for predicting subsequent success in the examinations. The multiple correlation is given in the last column and the regression coefficients in the preceding four columns.

(a) Pupils tested in 1951 at age 11.

Subject	N	M.H. Space Test 1	M.H. Intell. Test 46 (Verbal)	M.H. English Test 22b	M.H. Arith- metic Test 22	Multiple Corre- lation with Criterion
REGRESSION COEFFICIENTS						
Engineering Drawing (Intern- al exam., 1956) .....	47	+ .747	- .083	- .634	+ .301	+ .697
Engineering Drawing (G.C.E. Assoc. exam. Board, 1956)	27	+ .680	- .658	+ .501	+ .186	+ .582
Metalwork (G.C.E. Assoc. exam. Board, 1956) .....	21	+ .325	+ .003	- .364	+ .083	+ .336
Woodwork (G.C.E. Assoc. exam. Board, 1956) .....	23	- .004	- .328	+ .022	+ .071	+ .252
Metalwork (Theory and prac- tical. Internal exam., 1956)	39	- .426	- .163	- .666	- .048	+ .540

(b) Pupils tested in 1950 at age 13.

Subject	N	M.H. Space Test 1	M.H. Intell. Test 29 (Verbal)	M.H. English Test 12	M.H. Arith- metic Test 12	Multiple Corre- lation with Criterion
REGRESSION COEFFICIENTS						
Woodwork (Theory and prac- tical. Internal exam., 1953)	86	+ .656	- .367	- .426	- .223	+ .587
Woodwork (Practical only. Internal exam., 1953) ....	88	+ .650	- .190	- .374	- .229	+ .500

Mean multiple correlation = + .499.

In conclusion, some correlations with attainment in geometry are given for a class of 15-year-old grammar school pupils, though the data were not obtained in connection with the Middlesbrough investigation. The correlations, shown in Table 5, indicate the relationship between scores on Spatial Test I and the Manchester General Ability Test (Sen.) I<sup>22</sup> and scores on the Walton Geometry Attainment Test.<sup>19</sup> The latter is an objective test consisting of sixty-six geometrical problems, in which the information given for each problem is supplied in the form of an annotated geometrical diagram.

The correlations in Table 5 suggest that Spatial Test I (N.F.E.R.) is a better indicator of ability to solve geometrical problems (riders of the conventional type) than is an advanced verbal intelligence test of the type of Manchester General Ability Test (Sen.) I. Thus, there are grounds for supposing that Spatial Test I might have value for predicting success in courses in mathematics in grammar schools. In a recent N.F.E.R. publication,<sup>23</sup> Yates and Pidgeon have supplied evidence showing that Spatial Test I can make a significant contribution towards predicting grammar school success. The best prediction of success in the grammar schools included in the investigation (some 800 pupils) was given by the combination of scores  $4P.H.A. + 2E + Sp. I$  where P.H.A. is Primary Head Teacher's Assessment, E is a special English Test (by A. F. Watts) and Sp. I is Spatial Test I (by the writer). This weighted battery gave the best correlation with the criterion, namely .931. It is possible that this result may indicate that spatial ability is important for success in mathematical and scientific subjects.

TABLE 5

CORRELATIONS OF SCORES FOR TWELVE BOYS AND TWELVE GIRLS IN THE AGE-RANGE  
14:7 to 15:6.

(Class IVB in a Grammar School.)

	Spatial Test I (N.F.E.R.)	Manchester General Ability Test (Sen. I)	Geometry Attainment Test (Walton)
Spatial Test I (N.F.E.R.) .....	—	·569	·611
Manchester General Ability Test (Sen.) I .....	·569	—	·440
Geometry Attainment Test (Walton)	·611	·440	—

#### IV.—CONCLUSIONS.

A 'follow-up' over periods of three and five years has been carried out on groups of pupils who entered the Technical School in Middlesbrough in 1950 and 1951.

Correlations have been obtained suggesting that two spatial tests (N.F.E.R. Spatial Test I and M.H. Space Test I) give a better indication than other selection tests of subsequent performance in technical subjects such as engineering drawing, metalwork and woodwork. This was found to be true whether the criterion was based on internal examinations or on external G.C.E. examinations.

Regression coefficients and multiple correlations with the various criteria were also calculated. The regression coefficients for both spatial tests were in most cases substantial and positive, while more than half of the regression coefficients for the verbal intelligence tests were negative.

Evidence has also been obtained suggesting that Spatial Test I (N.F.E.R.) is a better measure than a verbal intelligence test of abilities which enter into attainment in geometry. This result was obtained with a fourth form of twenty-four pupils. It may partly account for the finding reported by Yates and Pidgeon that Spatial Test I, when used in conjunction with other tests, has been found to make a significant contribution towards the best prediction of all-round success in the grammar school.

ACKNOWLEDGEMENTS.—The writer is greatly indebted to Mr. T. H. C. Walker, Headmaster, Tynemouth Technical School, who was responsible for supplying most of the data in the first place; also to Mr. Stanley Hirst, Director of Education for Middlesbrough, who supplied information regarding the selection procedure employed in Middlesbrough in 1951 and who readily answered various queries.

Tribute must also be paid to the staff of Durham University Computing Laboratory and to one of the writer's former students, Mr. R. Y. Anderson, who assisted in various ways in connection with the calculation of correlation coefficients.

### V.—REFERENCES.

- <sup>1</sup> BAIN, J. T. (1946). *The Construction of a Space Test*. B.Ed. Thesis, Moray House, University of Edinburgh.
- <sup>2</sup> BRADFORD, E. J. G. (1948). Symposium on the selection of pupils for different types of secondary schools. IV—An experimenter's point of view. *Brit. J. Educ. Psychol.*, 18, 67-84.
- <sup>3</sup> DEMPSTER, J. J. B. (1950). *An Investigation into the Use of Verbal and Spatial Tests in the Selection of Boys for Grammar and Technical Schools*. Ph.D. Thesis, University of London.
- <sup>4</sup> DREW, L. J. (1947). The measurement of technical ability. *Occup. Psychol.*, 21, 34-48.
- <sup>5</sup> EMMETT, W. G. (1949). Evidence of a space factor at 11+ and earlier. *Brit. J. Psychol. (Stat. Sect.)*, 2, 3-16.
- <sup>6</sup> HOLLIDAY, F. (1940). The selection of apprentices for the engineering industry. *Occup. Psychol.*, 14, 69.
- <sup>7</sup> KOUSSY, A. A. H., El (1935). The visual perception of space. *Brit. J. Psychol., Monogr. Suppl.*, 20.
- <sup>8</sup> LAWLEY, D. N. (1943). The application of the maximum likelihood method to factor analysis. *Brit. J. Psychol.*, 33, 172-175.
- <sup>9</sup> MILLS, L. F. (1947). *The Properties of a Space Test*. B.Ed. Thesis, Moray House, University of Edinburgh.
- <sup>10</sup> PEEL, E. A. (1949). Evidence of a practical factor at the age of 11. *Brit. J. Educ. Psychol.*, 19, 1-15.
- <sup>11</sup> SHUTTLEWORTH, C. W. (1942). Tests of technical aptitude. *Occup. Psychol.*, 16, 175-182.
- <sup>12</sup> SLATER, P. (1940). Some group tests of spatial judgment or practical ability. *Occup. Psychol.*, 14, 40-55.
- <sup>13</sup> SLATER, P., and BENNETT, ELIZABETH (1943). The development of spatial judgment and its relation to some educational problems. *Occup. Psychol.*, 17, 139-155.
- <sup>14</sup> SMITH, I. MACFARLANE (1937). *The Form Perception Factor*. Ed.B. Thesis, University of Glasgow, 1937. (Summary in *Brit. J. Educ. Psychol.*, 3, 1938.)
- <sup>15</sup> SMITH, I. MACFARLANE (1948). Measuring spatial ability in school pupils. *Occup. Psychol.*, 22, 150-159.
- <sup>16</sup> SMITH, I. MACFARLANE (1950). *Spatial Test 1 and Manual*. Published for N.F.E.R. by Newnes.
- <sup>17</sup> SMITH, I. MACFARLANE (1952). *An Investigation into the Problem of Measuring Spatial Ability in Pupils of 11 to 12 and of Devising a Reliable Test for Selecting for Technical Education, together with a Study of the Personality Qualities of Pupils so Selected*. Ph.D. Thesis, University of Manchester.
- <sup>18</sup> SMITH, I. MACFARLANE (1954). The development of a spatial test. *Durham Res. Rev.*, 5, 19-33.
- <sup>19</sup> WALTON, R. D. (1947). *Geometry Attainment Test and Manual*. University of London Press.
- <sup>20</sup> WATTS, A. F., and SLATER, P. (1950). *The Allocation of Primary School Leavers to Courses of Secondary Education*. Published for N.F.E.R. by Newnes.
- <sup>21</sup> WATTS, A. F. (1953). A group performance test. *Bulletin of N.F.E.R.*, 2, 15-21.
- <sup>22</sup> WISEMAN, S., and COLLEAGUES (1960). *The Manchester General Ability Test (Sen.) 1*. University of London Press.
- <sup>23</sup> YATES, A., and PIDGEON, D. (1957). *Admission to Grammar Schools*. Published for N.F.E.R. by Newnes.



# SYMPOSIUM: CONTRIBUTIONS TO THE DIAGNOSIS AND AND REMEDIAL TREATMENT OF READING DIFFICULTIES

## I—THE INVESTIGATION OF READING PROBLEMS TODAY.\*

By M. D. VERNON

(*Department of Psychology, University of Reading.*)

**SUMMARY.** In spite of the large number of studies in recent years of problems connected with the processes of reading and with inability to learn to read normally, these problems are still very imperfectly understood. Although they are undoubtedly complex, failure to solve them appears to be due in part to the inadequacy of the methods employed in investigating them. Experimental studies of methods of teaching reading are often not suitably controlled, nor carried out for a sufficient length of time. The various factors associated with different types of reading backwardness are not always properly measured, nor their incidence and frequency assessed in different individual cases. The causal relationship between these factors and reading backwardness remains obscure. In part, these failures may be due to the fact that insufficient attention has been devoted to the nature of the reading processes and of their breakdown in backward readers.

### I.—INTRODUCTION.

THE problems of learning to read, and of the failure to learn to read normally, continue to attract extensive study. Yet, within the last few years there appear to have been few investigations which have added greatly to our understanding of these problems. We still know little as to the nature of the blockage which seems to prevent some children from learning to read normally, and particularly from acquiring the ability to recognize words. Nor do we know which are the processes which fail to develop. Some new evidence has been advanced as to the significant factors which characterize backward readers as contrasted with normal readers. But it is seldom clear to what extent these factors actually *cause* backwardness; or whether they are merely associated with it, and have no causal relationship with it.

Now it is inevitable that there should be considerable difficulty in establishing clearly principles as to the optimum methods of teaching reading, and the causes of failure, since the number of factors which affect learning to read is very large, and many of them are impossible to control experimentally. However, far more might be discovered if the experimental methods were better adapted to the type of study. But, although statistical techniques for the treatment of results have been refined and elaborated in recent years, there has been little improvement in the techniques of investigation. It may well be that it is for this reason that so little new light has been thrown on these problems.

### II.—READING SURVEYS.

No doubt, large-scale surveys of reading performance, such as that of Morris in Kent (1959), are of value to local education authorities in giving them some idea as to the reading achievement of children in their locality, and of the kind of conditions in which this is most and least satisfactory. In several recent

\* Based on a paper read to the South-Western Branch of the British Psychological Society in October, 1959.

surveys of this kind, the percentage of cases of severe reading backwardness has always appeared very small. Yet, in spite of these optimistic reports, teachers continue to complain about the children who do not learn to read normally, or fail to learn at all; and such cases are still frequently referred to educational psychologists and to clinics. Indeed, the Morris survey showed that about 20 per cent. of the 7-year-old children were still at the 'first primer' stage or below, meaning that they were poor readers or non-readers. No doubt, there is a large amount of comparatively temporary backwardness, much of which is relieved by suitable teaching; and, in some cases, the children may grow out of it themselves. We never learn, however, the amount of severe permanent backwardness. How many children, apart from low-grade defectives, never learn to read at all? And how many learn just enough to read simple notices and instructions, the betting news and the football pools, but cannot read a continuous text? This information should not be impossible to obtain; but until we have it, we cannot tell whether illiteracy is still a serious problem in this country; whether or not the current methods of teaching reading are about as effective as they can be; whether remedial teaching is overcoming such difficulties as normal school teaching cannot be expected to meet; and whether there is a considerable residue of adults and children who for some cause outside the teaching situation never learn to read normally.

### III.—TECHNIQUES FOR INVESTIGATING THE EFFECTS OF TEACHING METHODS.

Not only do methods of investigation in current use fail to disclose the amount of permanent illiteracy in the population; they also do not give us much reliable information as to the best methods of teaching reading and the probable causes of reading difficulty. Two main types of technique of investigation have been employed. The first, used for discovering the best methods of teaching reading, is the method of comparing the reading achievement of classes, either remedial or normal school classes, taught by different methods. In general, the results are likely to be more satisfactory if the children are specially selected and matched for intelligence, socio-economic status, etc., and the teachers and methods also chosen for the purpose. If already existent ordinary classes are studied, the number of variables relating to the children, the school conditions, etc., is so large as either to invalidate the results or make them impossibly complex to interpret. Thus, in Morris's survey, it appeared at first sight that children in schools where phonic teaching was introduced at a very early age, showed a better reading achievement in the primary school than children from schools in which phonics were introduced later. But this difference was significant only when allowance was made for the fact that the non-verbal intelligence of the children in the former schools was lower than that of the children in the latter. This somewhat surprising finding makes it difficult to decide whether or not the phonic methods were really superior. On the other hand, until the effects of intelligence were partialled out, it appeared that size of school, urban locality and socio-economic status all correlated positively with reading achievement; but this result must be interpreted as showing only that more intelligent children, who have a higher average reading achievement, are found in better-class urban neighbourhoods where there are large schools. (It is only fair to note, incidentally, that Morris has since been carrying out intensive studies of various factors in classroom teaching, such as teacher training, which, it appears, may be of considerably more importance than the use of phonic or look-and-say in enabling children to learn to read without difficulty.)

It is perhaps worth comparing these findings with those of Kemp (1955) which gave rise to some comment a few years ago. His method was to test the

achievement in reading and in other school subjects of a group of thirty 11-year-old children from each of fifty mixed junior schools in the London area. There was no special design as to teachers and teaching methods. Here again, the relationships between children's intelligence and socio-economic status, and the size of school and class, seemed to be so inextricably intermingled that it was difficult to say if any of these factors, apart from intelligence, had any effect on reading achievement.

Much clearer results are obtained if some attempt is made to control these variables, and to measure the achievement of small groups of children specially selected, and with teaching methods as far as possible controlled. A good example of such an enquiry was that of Daniels and Diack (1956) into the efficacy of their modified phonic method of teaching backward 8-year-olds to read, as compared with the efficacy of a 'mixed' method. But, even in this enquiry, which appeared to show an undoubted superiority for the phonic method, the competence of the teachers of the two types of class was not necessarily the same, thus introducing an important variable. Only by repeating such an experiment many times with teachers of varying competence in both types of class could this variable be compensated for. Moreover, some conclusion might then be possible as to the ability of teachers with varying degrees of competence to utilize the various teaching methods.

Ace (1956) attempted to get over the difficulty of varying competence by having the same teacher give remedial reading teaching to two groups, one by means of the 'Moxon' method and the other by 'mixed' methods. The former gave the better results. But even if this teacher was able to apply both methods with the same degree of competence and enthusiasm, it is still not justifiable to generalize from the findings that remedial teachers in general would obtain better results with the 'Moxon' method.

The efficiency of the various methods is commonly assessed by comparing the average test performances of the children taught by them. Valuable supplementary evidence could be obtained, however, by discovering how many failed to profit by them, and continued to show an appreciable degree of backwardness. Indeed, it is probably more important to know whether a method is generally fairly successful (i.e., results in few cases of backwardness) than whether it produces a slightly higher average achievement.

Another criticism which must be advanced against experiments of this kind is that they are seldom carried out for a sufficient period; and the same is true with regard to the success of remedial teaching in general. We are, for instance, informed that at the end of three months or six months, the children taught by one method have improved their reading to a greater extent on the average than those taught by another method. Now Sparks and Fay (1957) found that a group of children taught to read by means of a method called 'Phonetic Keys to Reading' made more rapid progress at first than a group taught by the ordinary method, which contained less phonics. By the end of the fourth grade, however, there was no significant difference in achievement between the two groups. But more positive results were obtained by means of a follow-up in the well-known enquiry of Gardner (1950) which compared the effect of 'free expression' methods with that of formal methods in the infant school. The follow-up, when the children were 9 or 10 years old, showed an advantage for the 'free expression' methods, which it was said produced no reading failures and a higher average reading achievement. But these findings were probably affected by differences in the competences and enthusiasm of the respective teachers.



Again, we may be informed that the effect of a three or six-month period of remedial reading teaching has been to raise the children's average reading age by a year or more. But we are not justified in concluding that the same improvement is maintained indefinitely, and that the children eventually become normal readers. Thus, Collins (1956) compared the improvement in word recognition of three matched groups of backward readers: (1) attending a remedial education, centre; (2) given remedial teaching in school; (3) a control group. At the end of a six-month period, the first group showed a significantly greater improvement than the second and third groups, who were not significantly different from each other. But follow-up tests two or three years later showed no difference between the groups; the first group had slowed down and the other two had caught them up. There is, indeed, evidence from other sources that the long-term effect of remedial reading teaching may be rather slight, unless the teaching is greatly prolonged.

There are certain circumstances in which, however, a comparatively brief follow-up may give useful results. Roslow (1940) studied the effect of 'reading readiness' activities in the kindergarten on the progress in reading of children when they entered Grade I of the ordinary school. The children were at this point divided into three groups according to their mental age and their aptitude for reading as assessed by 'reading readiness' tests; and they were then taught along lines indicated by their performance on the tests. It was found that the majority of them made satisfactory progress during their year in Grade I; and it could be assumed with some justification that they would continue to do so. Another comparatively short-term study was made by Mitzel and Medley (1957), of the relative degree of improvement shown by children taught by different teachers over a six-month period. It was found that the average differences between children in different classes in the same school were almost as great as those between children in different schools. This finding is valuable in that it shows the effects of different degrees of competence in the teachers over quite a short period of time. However, it would also be interesting to know whether the children who lagged behind through incompetent teaching caught up again if they were subsequently taught by better teachers.

#### IV.—STUDIES OF GROUPS OF RETARDED READERS.

We must now consider the reliability of the method most commonly used for investigating the causes of failure in learning to read—namely, the selection of a group of children retarded in reading and the comparison of their characteristics with those of a group of normal or superior readers. Numerous criticisms may be made of the procedure adopted in many of these investigations. In the first place, the degree and nature of the backwardness may be very variable. Often the children are said to be one year or a year-and-a-half below the normal reading test performance for their chronological age, their mental age or even their 'grade.' The last is a particularly unsatisfactory measure, since mental age within a grade may vary by five years or more. Even if reading performance is related to mental age, this mental age differs according as to whether intelligence has been tested by a group written test, an orally given verbal test or a non-verbal test. Obviously, the first of these is quite unsatisfactory for retarded readers, and, indeed, impossible for non-readers. But it has been argued that the non-verbal tests are unsatisfactory also in that they give no measure of the child's general ability to use language. Probably the most satisfactory assessment is given by the WISC, which measures both oral verbal and non-verbal

intelligence, and thus affords a measure of retardation below both the verbal and the non-verbal M.A. But individual intelligence tests are rejected by many experimenters as being too time-consuming.

However, the criterion of retardation as a performance one year or a year-and-a-half below the standard reading test performance is not in any case a good one. It represents an attempt to quantify on a single scale performances which may, in fact, be qualitatively different from one another: that of the child who cannot read at all, or can merely recognize a few words he has learnt by heart; and of the child who knows how to read new words, but perhaps reads only one word at a time, and fails to understand their meaning. The latter type of backwardness, tiresome as it may be in school, should be considered as falling into a different category from the former. It is possible that the underlying causes of these conditions are different; but this we cannot tell until a clear distinction is made between their incidence.

When experimenters endeavour to judge the effect of various causal factors in producing reading backwardness, they may find some difficulty in deciding on the best methods of assessing these factors and their association with the reading backwardness. The usual practice with discontinuous functions, such as laterality, is to compare the frequency of occurrence of these in comparable normal and backward readers by means of a contingency coefficient. But the urge is strong to convert such measures into indices which can be submitted to more elaborate parametric statistics. Thus, even with laterality some experimenters have concocted an index,  $\frac{R}{R+L}$  or  $\frac{R-L^*}{R+L}$ , which is supposed to

measure degree of laterality. With other less obviously discontinuous functions, the tendency is stronger. There is no *a priori* reason why actual test scores, say of perceptual span, should not be correlated with scores on reading tests; or converted into averages, and the significance of the difference of these calculated as between backward and normal readers. But there is, nevertheless, the considerable disadvantage in these procedures that we cannot tell whether *all* the backward readers are to some extent deficient in these functions; or whether some are highly deficient and others normal, which is much more probable. Therefore, even in these cases, it may be better to determine the frequencies of low scores, and hence demonstrate how characteristic this defect is of backward readers.

The calculation of averages or correlations of grades on various qualities which cannot be scored numerically appears not only to suffer from the disadvantage mentioned, but also to be statistically suspect. It might be better to submit these to some form of non-parametric statistics, since they are neither continuous nor linear functions. This point may be illustrated by reference to the recently published monograph by Malmquist (1958), which describes a very detailed and far-reaching study of the reading performance of 400 first-grade Swedish children. These children were divided on the basis of reading test scores into three groups: poor readers, who scored 1σ or more below the norm; good readers, who scored 1σ or more above the norm; and medium readers, from +1 to -1σ. Incidentally, very few of these children appear to have been non-readers; perhaps the Swedish language is comparatively easy to learn to read, in the early stages, and there are very few non-readers. Malmquist determined the frequency in the three groups of a number of factors such as: social group, parents' education and other social indices (significant association

\* R and L being the number of activities in which the right or left hand, foot or eye is preferred, respectively.

with poor reading); prematurity of birth (not significant); speech defects (significant); visual defects (not significant); laterality (not significant); pre-school diseases (not significant). She also used this method for certain personality factors, such as nervous symptoms reported by parents during the pre-school period, and found the latter to be more frequent for the good than the poor readers. But for personality qualities shown in school, such as concentration, persistence, self-confidence, dominance-submissiveness, stability-nervousness, she obtained ratings from the teachers on a five-point scale, correlated these ratings with reading scores, and also calculated the significance of the differences between the average ratings for the three groups. These were said to be significantly different, and the correlations were of the order of 0.3-0.4.

These methods are some of the most frequently used in reading studies; but what do they prove to us? There is somewhat less of these desirable qualities among the poorer than among the better readers. But to what extent is each of these qualities defective in how many backward readers? Furthermore, were lack of concentration, persistence and emotional stability the cause or the effect of difficulty in learning to read? Among older children particularly, many of the symptoms of emotional maladjustment may be caused by their failure in reading. In first-grade children it should be possible to distinguish between cause and effect. Yet apparently among Malmquist's children, the poor readers had fewer nervous symptoms before coming to school, but were less stable in school, once they had begun learning to read. If the evidence of the teachers' ratings can be accepted, it would seem that reading backwardness had produced this loss of stability.

Another instance of the difficulty of distinguishing cause and effect is given by the frequently quoted association between visual perceptual ability and reading achievement. Both Malmquist (1958) and Goins (1958) found that among first-grade children good readers performed better than poor readers on tests of perception. But the superiority seems to have been most apparent in the perception of digits and letters, in Malmquist's experiments; and in any case, was largely a function of superiority in intelligence. In Goins' experiments the superiority was in pattern copying and absence of reversals in perception. But may not these superior performances have been due, at least in part, to the reading (and writing) achievement of the better readers, which had taught them the necessity of distinguishing small details accurately? For when Goins went on to determine the effect on reading of practice in the tachistoscopic perception of figures, digits, letters and words, she found that, although tachistoscopic perception improved considerably, the effect on reading was inconsiderable.

In another recent survey of the possible causes of reading backwardness, there has been a real attempt to determine not only the association between certain factors and backwardness, but also which of these factors may be causal. Stott (1959) examined the records of illness in early childhood, the personality characteristics and the reading backwardness of an unselected sample of 142 children who had spent at least 2 weeks in hospital during infancy. He also compared them with a control group of children of the same age and sex who were not known to have been in hospital. The hospitalized children were more backward than the controls, and their backwardness was also associated with the number and type of diseases from which they had suffered, and with a personality characteristic termed 'unforthcomingness,' which seems to denote an impairment of motivation and inability to attempt actively to overcome difficulties. In an earlier study (1957), Stott had established an association between maternal stress during pregnancy, frequent early illness and low intelligence; and between backwardness in school and 'unforthcomingness'. The



general conclusion from the present survey was that stress during pregnancy might set up a general syndrome, which could be termed 'weak constitution,' giving rise to frequent illness, mental retardation and unforthcomingness; but that it was also possible that the retardation might be enhanced, if not caused directly, by defective motivation, which, in turn, might have resulted from the early illness. The factors of social class difference, and of separation from the mother in infancy by hospitalization, were found not to be of importance.

Clearly much further study is required to define the exact effects of these various factors, though undoubtedly the greatest impairment may be produced by a multiplicity of causes. It is often difficult to follow Stott's reasoning as to which was cause and which effect. Nor did he take into account the existence of 'maternal over-protection' which may very well affect children who have suffered much illness in early childhood, and which almost certainly tends to aggravate the personality quality he terms 'unforthcomingness.' Again, he appears to equate poor reading achievement, general educational backwardness and low intelligence. Though some reading backwardness may be associated with low intelligence, we know that some of the most severe cases are of normal intelligence. But Stott's measurement of reading achievement was too rough to pick out such cases. Thus, it may well be that they are quite distinct from the cases of general subnormality studied by Stott.

There are also cases in which it appears that there is unlikely to be any direct association between a certain factor and reading achievement; but that the association is with some other factor or factors not measured by the experimenter. This seems to have been the case in the study carried out by Lynn (1955), which appeared to demonstrate that anxiety might improve reading rather than impairing it. Eighty junior school children, aged  $7\frac{1}{2}$ -11 years, selected to include good, average and poor readers, were asked questions drawn from the Taylor Anxiety Scale about personal anxieties such as fear of the mother's death, dislike of being alone, etc., and questions about impersonal anxieties such as fear of animals, thunder, nightmares, etc. Significant correlations were obtained of 0.33 between reading attainment and personal anxieties, and of 0.24 with impersonal anxieties. In another paper (1957), Lynn suggested that anxious children 'withdraw' into reading, and hence showed high achievement in it. On the other hand, they were relatively poor at arithmetic, usually considered to be difficult for emotionally unstable children. Now other studies have shown that some emotionally maladjusted children who were good readers used reading either as a phantasy refuge from the real world, or as a compensation for their other deficiencies. It may also be that some of the better readers in Lynn's study were over-anxious to succeed in the 11+ examination. It would, however, be extremely unwise to conclude that anxious children in general are good readers, or that their anxiety causes their superiority in reading. It seems far more probable that both reading achievement and anxiety were correlated with some other quality, such as ambition.

#### V.—INDIVIDUAL STUDIES.

It seems possible that it is only by the method of individual study that we may eventually be able to demonstrate the factors or patterns of factors which are significant in cases of reading difficulty. From such experimental and clinical evidence as we have, it appears almost certain that causation is multifactorial, and that different factors operate in different cases. Thus, a recent

study by Spache (1957) using the Rosenweig Picture Frustration test showed that the responses of backward readers (of various ages) could be grouped into those of :

- (1) The overtly hostile, aggressive and intolerant.
- (2) Those below normal in hostility, excessively apt to accept blame.
- (3) The defensive, hyper-sensitive to criticism.
- (4) Those who tended to withdraw in the face of difficulties.
- (5) Those who would not accept failure, but drove themselves on in spite of all obstacles.

These observations, in so far as conclusions based on a single projection test can be accepted as valid, demonstrate patterns of personality which could be looked for in individual studies also ; and if they appeared, could be related to environmental factors such as parental treatment. They may, of course, have been caused by reading failure in some cases, rather than having caused it ; but adequate case histories and early testing might throw some light on this.

One method of clasifying the data from individual cases is to record in tables the presence or absence of relevant factors in each case. It is interesting to note that Malmquist (1958) adopted this procedure with thirty-four of her children of normal or above normal I.Q. whose reading performance was markedly inferior to that expected for their M.A. But she made no further use of these data. From a casual inspection, these cases appear to have come mainly from families of low socio-economic status, the parents having themselves had little education ; they suffered from few speech or other defects, but in many cases had had severe or prolonged illnesses before coming to school ; on the whole they possessed normal sociability, but many were lacking in persistence, concentration and emotional stability. This suggests a picture somewhat resembling Stott's of the impoverished, uncultured and physically sub-normal child who often finds difficulty in starting learning and tackling difficult tasks when he comes to school ; but who, in many cases, has little prolonged or severe reading disability. As I mentioned above, there seem to have been few of the latter cases in Malmquist's sample. A tabulation of severe cases—non-readers, for instance—would probably show different patterns of symptoms and causes. It would, of course, require a much more detailed diagnosis of personality qualities and characteristics of the home background than that obtained by Malmquist.

Little evidence has been brought forward in recent years as to the incidence or nature of the disorder sometimes called 'specific dyslexia,' characterized in some cases, though not in all, by language deficiencies, motor inco-ordination, incomplete laterality, hyperactivity and some emotional instability. A study by De Hirsch (1954) of children of this type showed that they may also be deficient in figure-ground differentiation and in the ability to form abstract concepts—though to a less extent than are cerebral-palsied children. This suggests some form of neurological deficiency in the 'specific dyslexic.' But positive evidence corroborating this hypothesis is lacking, and is, of course, very difficult to obtain.

## VI.—CONCLUSIONS.

It may well be that no definite evidence as to the causes of reading disability will be forthcoming until more is known as to the fundamental nature of the disability—exactly what it is that these children cannot do. At present, it appears that there must exist some failure in reasoning related to the use of

language which precludes these children from analysing the printed words, systematically associating sounds to the constituent letters and synthesizing these to form the total word sound. But much more detailed experimental study of the processes of reasoning in backward readers is necessary to elucidate the exact nature of the failure.

#### VII.—REFERENCES.

- ACE, P. W. (1956). A remedial teaching scheme: introducing a new reading method. *Brit. J. Educ. Psychol.*, 26, 191-193.
- COLLINS, J. E. (1956). *The Effects of Remedial Education*. Ph.D. Thesis, University of Birmingham.
- DANIELS, J. C., and DIACK, H. (1956). *Progress in Reading*. University of Nottingham Institute of Education.
- DE HIRSCH, K. (1954). Gestalt psychology as applied to language disturbances. *J. Nerv. Ment. Dis.*, 120, 257-261.
- GARDNER, D. E. M. (1950). *Long-term Results of Infant School Methods*. London: Methuen.
- GOINS, J. T. (1958). Visual perceptual abilities and early reading progress. *Suppl. Educ. Monogs.*, No. 87.
- KEMP, L. C. D. (1955). Environmental and other characteristics determining attainment in primary schools. *Brit. J. Educ. Psychol.*, 25, 67-77.
- LYNN, R. (1955). Personality factors in reading achievement. *Proc. Roy. Soc. Med.*, 48, 996-998.
- LYNN, R. (1957). Temperamental characteristics related to disparity of attainment in reading and arithmetic. *Brit. J. Educ. Psychol.*, 27, 62-67.
- MALMQUIST, E. (1958). *Factors Related to Reading Disabilities in the First Grade of the Elementary School*. Stockholm: Almqvist and Wiksell.
- MITZEL, H. E., and MEDLEY, D. M. (1957). Pupil growth in reading—an index of effective teaching. *J. Educ. Psychol.*, 48, 227-239.
- MORRIS, J. M. (1959). *Reading in the Primary School*. London: Newnes.
- ROSLOW, S. (1940). Reading readiness and reading achievement in the first grade. *J. Exper. Educ.*, 9, 154-159.
- SPACHE, G. (1957). Personality patterns of retarded readers. *J. Educ. Res.*, 50, 462-469.
- SPARKS, P. E., and FAY, L. C. (1957). An evaluation of two methods of teaching reading. *Elemen. School J.*, 57, 386-390.
- STOTT, D. H. (1957). Physical and mental handicaps following a disturbed pregnancy. *Lancet*, ii, 1006.
- STOTT, D. H. (1959). Infantile illness and subsequent mental and emotional development. *J. Genet. Psychol.*, 94, 233-251.

(Manuscript received, December 2nd, 1959.)



## II—THE EFFECT OF PRACTICE ON PERFORMANCE IN SCHOLASTIC TESTS.

BY W. CURR AND N. GOURLAY

(*Education Department, University of Birmingham.*)

SUMMARY. 1.—The main object of the investigation was to obtain some measure of practice effect for tests in mechanical reading, reading comprehension and arithmetic. Standardised tests in these skills were given at varying intervals of time to groups of 9–10-year-old children in four primary schools.

2.—The statistical analysis employed provided estimates of the mean gains both at the average level of ability for the age group (9·5 years) and two years below average (7·5 years).

Practice effect was greatest for reading comprehension. At the 9·5 level, a re-test after one month produced a mean gain of 10·1 months and further re-tests after 3 and 6 months gave gains of 18·2 and 26·9 months, respectively. The corresponding figures at the 7·5 level were 8·9, 10·4 and 12·2 months.

For the mechanical reading test, the gains were 5·3, 8·0 and 14·1 months at the 9·5 level and 4·4, 6·2 and 11·8 at the 7·5 level. Practice effect was only slight for the arithmetic test.

3.—Practice effect was also estimated for other cases. In particular, pupils tested on both forms of the reading comprehension test at an interval of one month registered gains of 7·2 and 4·2 months at the 9·5 and 7·5 levels, respectively.

4.—Reference was made to our earlier investigation into the scholastic results of remedial education. It was pointed out that, in the light of the data here presented, the gains made by the controls in our experiment could be explained entirely in terms of maturation and practice effect. Data are also given to show the impermanence of the net gains made by the pupils receiving remedial education. Corroborative data from another investigation are also presented.

### I.—INTRODUCTION.

THIS investigation is a sequel to an experiment in remedial education reported a few years ago in this *Journal*.<sup>1</sup> Earlier attempts at assessing the scholastic results of remedial education had used a simple test-retest procedure. We maintained that *gross* gains, as measured by tests, were inflated by a number of factors, in particular (a) practice effect; (b) regression effect; and (c) a coaching effect; and that these factors must be allowed for before a valid assessment of the efficacy of remedial education can be obtained.

In our experiment, we allowed for the first two factors by introducing a control group of pupils who received no special instruction but who were subjected to exactly the same testing programme as the remedial group. The 'true' gain of the remedial group was then defined as the amount by which the gross gain of the remedial group exceeded the gross gain of the controls. (We attempted to demonstrate the coaching effect by showing that the 'true' gain for reading varied with the type of reading test used).

From the standpoint of the remedial educationist, the results of our experiment were disappointing. The *gross* gains of the remedial group were comparable with the gains reported for other investigations but, since appreciable gains were also made by the controls, the 'true' gains were small. It was only in reading, as measured by a Schonell Reading Comprehension test, that a statistically significant 'true' gain was obtained.

Obviously, the results of our experiment would have been more favourable to remedial education if the gross gains of the controls had not been so large. The suggestion was, therefore, put forward by Birch<sup>2</sup> and supported later by Pringle and Gulliford<sup>3</sup> that either the class teachers had paid special attention to the control pupils or they had been so influenced by the experiment that they were able to bring about substantial over-all improvement in their class work. We, of course, favoured the simple explanation that the gains of the controls had been inflated mainly owing to the effect of practice in the tests used but, at the time, we had no practice effect data to confirm this.

The present investigation was, therefore, carried out in order to obtain some measure of the amount of practice effect which can occur when the common types of scholastic test are given at varying intervals of time. Also, in view of the likelihood of differential practice effects, it was intended that the investigation should provide a measure of practice effect not only at the average level, but at the level of the pupils to be found in remedial groups.

A section on coaching appears at the end of the paper in which, as a result of further data, we broaden the conception of coaching given in our earlier study. The section has been added in order to indicate more fully our present position in regard to remedial education.

## II.—PLAN OF INVESTIGATION.

(i) *Selection of subjects.*—The entire testing programme was carried out in four Birmingham primary schools by four members of the Diploma course in the Psychology of Childhood, each operating in the one school throughout the investigation.\* In our remedial education experiment, we had selected children between 8 and 9 years of age, but as many of the children in this group scored zero in one or two of the tests, particularly the Reading Comprehension test, it was decided that for the present investigation, children between 9 and 10 years only would be included. A further reason for this change was that most other investigations in this field tend to deal with older children, on the grounds that they are more likely to present true cases of underfunctioning, as distinct from mere discrepancies in attainment arising from inequalities in rate of development.

In each of the four schools, a list was drawn up of the names of all the children between 9 and 10 years of age. From each list, four groups—each consisting of six boys and six girls—and two groups—each consisting of twelve boys and twelve girls—were chosen at random. (We used random numbers for this purpose).

In the account which follows the four schools are denoted by the letters P, Q, R, S, and the groups by the letters A, B, C, D, and E, F, respectively.

(ii) *Tests.* The following Schonell tests were used :

- (a) Reading Comprehension test, Forms A and B (RCA and RCB).
- (b) Graded Word Reading test (GW).
- (c) Mechanical Arithmetic test, Form A (MA).

For convenience, we shall sometimes refer to these tests by means of the abbreviations in the brackets.

\* Some doubt has been expressed concerning the reliability of test scores when the tests are administered by Diploma students. An investigation, which we carried out recently,<sup>4</sup> shows that even in the case of administering the more difficult Terman-Merrill test, Diploma students are little short of the standard of fully qualified psychologists.

In the beginning a spelling test was also included but had to be dropped because of the limited amount of time available for testing. However, a small amount of data was obtained for the spelling test which suggested that, as with the arithmetic test, practice effect is not very great. (The data are not reproduced here).

(iii) *The testing programme.* The testing programme in each school was as follows :

Tests	Groups tested			
	Nov.	Dec.	Mar.	June.
RCA	E	F	—	—
RCB	A F	AB E	AC	ABCD
GW	A	AB	AC	ABCD
MA	A	AB	AC	ABCD

It will be seen that the A-groups were tested in November and re-tested on each subsequent occasion of testing, i.e., at intervals of 1 month, 3 months and 3 months ; the B-groups in December and June only ; and so on.

### III.—RESULTS.

(i) *Mean gains (Groups A, B, C and D).* Tables 1, 2 and 3 show the initial mean scores and the subsequent gains (as measured from initial score) for the A, B, C and D groups in each school. (There was no evidence of sex differences. The results for boys and girls were, therefore, combined).

TABLE 1

INITIAL MEAN SCORES AND GAINS (IN MONTHS) ACHIEVED BY A, B, C and D-GROUPS IN SCHONELL'S READING COMPREHENSION TEST, FORM B.

School	Group A				Group B		Group C		Group D
	Initial score (Nov.)	Gain			Initial score (Dec.)	Gain (Jun.)	Initial score (Mar.)	Gain (Jun.)	Score (Jun.)
		Dec.	Mar.	Jun.					
P	101.9	9.6	11.6	17.0	110.2	5.1	100.6	9.5	96.2
Q	106.1	11.9	19.0	22.1	108.0	8.5	107.0	10.0	115.1
R	108.7	8.5	16.4	26.2	114.0	16.0	112.5	14.1	123.6
S	91.7	3.7	9.0	9.9	88.3	7.5	102.0	5.8	94.0
Total	102.1	8.4	14.0	18.8	105.1	9.3	105.4	9.9	107.2



TABLE 2

INITIAL MEAN SCORES AND GAINS (IN MONTHS) ACHIEVED BY A, B, C AND D-GROUPS IN SCHONELL'S GRADED WORD READING TEST.

School	Group A				Group B		Group C		Group D
	Initial score (Nov.)	Gain			Initial score (Dec.)	Gain (Jun.)	Initial score (Mar.)	Gain (Jun.)	Score (Jun.)
		Dec.	Mar.	Jun.					
P	113.2	0.7	8.3	12.7	118.4	11.0	109.0	5.0	102.2
Q	116.0	9.5	13.6	18.9	119.8	5.7	122.2	3.7	124.0
R	113.2	4.8	4.4	9.5	120.2	6.9	126.7	7.0	128.2
S	84.4	2.9	5.3	11.0	75.5	13.9	97.2	5.6	89.0
Total .....	106.7	4.5	7.9	13.0	108.4	9.4	113.8	5.3	110.9

TABLE 3

INITIAL MEAN SCORES AND GAINS (IN MONTHS) ACHIEVED BY A, B, C AND D-GROUPS IN SCHONELL'S MECHANICAL ARITHMETIC TEST FORM A.

School	Group A				Group B		Group C		Group D
	Initial score (Nov.)	Gain			Initial score (Dec.)	Gain (Jun.)	Initial score (Mar.)	Gain (Jun.)	Score (Jun.)
		Dec.	Mar.	Jun.					
P	98.9	3.8	8.9	8.5	102.0	5.6	101.8	2.4	97.7
Q	104.9	4.5	8.3	12.8	107.2	14.7	115.0	4.2	129.8
R	113.3	1.6	6.4	5.3	118.2	6.2	121.6	1.3	128.1
S	97.9	1.7	4.1	9.1	91.6	4.6	107.4	3.9	109.8
Total .....	103.8	2.9	6.9	8.9	104.8	7.8	111.5	3.0	116.4

Several points can be made :

(a) The schools differ significantly in regard to their initial mean scores (as was shown by an analysis of variance).

(b) The differences between schools in regard to mean gains are also significant.\* Nor do these differences disappear when allowance is made for the differences in level of the schools (as was shown by an analysis of covariance).

(c) Standard errors can be calculated for each of the gains shown in Tables 1, 2 and 3. But, in assigning a standard error to any one gain, it is necessary to know what variation is being taken into account.

The gain for any one pupil can be analysed into (i) a school component (the mean gain of the school to which the individual belongs); (ii) an individual component (which may be positive or negative); and (iii) a random component (to explain fluctuations in the gain of the individual from testing to testing).

If group mean gains are to be taken as estimates of the mean gains for the schools (with respect to which the groups are only samples), standard errors must allow for variation due to (ii) and (iii). For the three mean gains of group A, standard errors were found to be of the order 2, 2.5 and 3 months, respectively (for all three tests) and for the mean gains of groups B and C of the order 2.5 and 3 months, respectively.

We come now to the question of the standard errors of the mean gains derived from the four schools taken together. As estimates of the mean gains for the four schools only, their standard errors are about half the size of the standard errors already given. But as estimates of the mean gains for the total population of schools, of which our four schools are only a sample, their standard errors are very much larger since allowance must now be made for variation due to (i) as well as (ii) and (iii). However, our figures at least give an indication of the range of gains to be expected from school groups. To have attempted more would have meant testing children in a large number of schools, a task which was beyond our resources.

(d) There remains the matter of interpreting the gains. An examination of the magnitude of the gains, particularly the gains made by the A-groups at their second testing (one month after the first), shows that they cannot be attributed entirely to maturation. (Here we are using the term to denote the genuine improvement in performance which we would expect the children to make over the period of testing). Obviously, practice effect is present.†

If we assume a normal rate of maturation—one month of development per month of time—it will be seen that the effect of practice is about twice as great for the Reading Comprehension test as for the Graded Word Reading test. In the case of the Arithmetic test, practice effect would appear to be slight.

It was intended that the design of our investigation would provide us with a measure of the mean rate of maturation of the four schools. For the overall mean scores obtained by the A, B, C and D-groups, when they were first tested, can be taken as estimates of the mean maturation level of the schools for November, December, March, and June, respectively. These mean scores are reproduced in Table 4.

\* In his criticism of our earlier investigation, Birch regarded the significance of the schools variance as evidence that remedial education was being effective in some schools if not in others. But this could only have followed from a significant interaction variance (treatments  $\times$  schools).

† Other factors must also have operated (e.g., variations in conditions at the time of testing) but as they are random we shall ignore them in our discussion of the interpretation of the gains. It is to be noted that the standard errors given in the previous section are underestimates in that they do not take account of this source of variation.

TABLE 4

OVERALL MEAN SCORES OBTAINED BY A, B, C and D GROUPS WHEN TESTED FOR THE FIRST TIME.

Test	Mean Scores			
	A-group (tested Nov.)	B-group (tested Dec.)	C-group (tested Mar.)	D-group (tested Jun.)
RCB	102.1	105.1	105.4	107.2
GW	106.7	108.4	113.8	110.9
MA	103.8	104.8	111.5	116.4

The figures in the table are obviously too irregular to give a good indication of the true rate of maturation for the four schools. Their standard error works out at 3 months (and is probably higher if we take into account the possibility of other random factors being present such as variations in testing conditions). Even so, the figures can be used to establish the fact that maturation alone is not sufficient to explain the final gains made by the A-groups in the two reading tests. An analysis of variance comparing the final scores of the A-groups (obtained at the June testing) with the scores of the D-groups (also tested in June) produces an F-ratio around the 5 per cent. level of significance. In the case of the Arithmetic test, the mean score of the D-groups exceeds the final mean score of the A-groups, but the difference is, of course, not statistically significant.

(ii) *Differential Practice Effect (Groups A, B, C and D).* It is generally accepted that bright children benefit more from practice in a test than dull children. This fact has been demonstrated in the case of intelligence tests by Peel<sup>5</sup> and by Wiseman and Wrigley.<sup>6</sup> Actually, Peel showed that practice effect increased up to a certain point (slightly more than one standard deviation above the average) and then fell off; but this was not confirmed by Wiseman and Wrigley who found a steady increase in practice effect with level of ability.

Since a differential practice effect is almost certain to operate in the case of scholastic tests, it follows that a simple average of the mean practice effects for our four schools is of limited value. It is not the best estimate which we can obtain from our data of practice effect for the average 9-10-year-old since it makes no allowance for the differences in level of ability of the four schools. Furthermore, we wish to obtain a measure of practice effect not only at the average level of ability but at some point about two years lower.

Our method for obtaining the differential practice effects is described in section A of the Appendix. (Actually, the method gives differential gains involving both maturation and practice effect. As in the previous sub-section, an allowance has to be made for maturation in order to assess the magnitude of practice effect only.)

Because of school differences, the data for each school had again to be treated separately before any overall averages could be taken. The gains at the average level (9.5 years) and two years below average (7.5 years) were derived for the two reading tests and are shown in Tables 5 and 6. (Since practice effect was small for the arithmetic test, an analysis of the data for this test was not considered to be worth the labour involved).



TABLE 5

DIFFERENTIAL GAINS (IN MONTHS) FOR READING COMPREHENSION TEST.

School	Reading level	Group A			Group B	Group C
		Gain			Gain	Gain
		Nov.-Dec.	Nov.-Mar.	Nov.-Jun.	Dec.-Jun.	Mar.-Jun.
P	9.5 yrs.	8.9	13.0	23.6	5.9	6.6
	7.5 yrs.	12.6	10.6	13.2	5.1	10.6
Q	9.5 yrs.	13.0	22.0	26.9	7.6	12.9
	7.5 yrs.	9.7	12.9	11.1	10.7	7.0
R	9.5 yrs.	8.5	18.1	29.2	15.4	18.1
	7.5 yrs.	8.5	8.6	13.0	18.3	11.8
S	9.5 yrs.	9.7	19.6	27.8	22.7	13.8
	7.5 yrs.	4.6	9.4	11.6	13.3	-1.9
Total ....	9.5 yrs.	10.1	18.2	26.9	12.9	12.9
	7.5 yrs.	8.9	10.4	12.2	11.9	6.9

TABLE 6

DIFFERENTIAL GAINS (IN MONTHS) FOR GRADED WORD READING TEST.

School	Reading level	Group A			Group B	Group C
		Gain			Gain	Gain
		Nov.-Dec.	Nov.-Mar.	Nov.-Jun.	Dec.-Jun.	Mar.-Jun.
P	9.5 yrs.	0.5	8.5	12.7	10.3	5.6
	7.5 yrs.	2.0	7.4	11.9	7.1	4.1
Q	9.5 yrs.	9.0	13.1	18.8	5.3	4.1
	7.5 yrs.	7.3	7.2	15.8	4.1	5.4
R	9.5 yrs.	4.8	4.3	9.6	7.2	7.0
	7.5 yrs.	4.7	4.1	7.4	9.0	7.1
S	9.5 yrs.	6.8	6.1	15.4	21.0	8.0
	7.5 yrs.	3.6	5.9	11.9	16.6	5.5
Total ....	9.5 yrs.	5.3	8.0	14.1	11.0	6.2
	7.5 yrs.	4.4	6.2	11.8	9.2	5.5

It will be seen that, on the whole, practice effect at the 7.5 year level is not so much less than it is at the 9.5 year level. The big difference occurs with the A-groups when re-tested on the Reading Comprehension test. Whereas the pupils at the 7.5 level show no practice effect after the first re-testing, the pupils at the 9.5 level show considerable practice effect with each re-testing. Probably one reason why this occurs with the RC test and not with the GW test is that the former, unlike the latter, has a time limit and this is likely to be to the advantage of the brighter pupil at the second and third re-tests.

(iii) *Practice effect (Groups E and F).* The difficulty in estimating practice effect in going from one form to the other of the Reading Comprehension test is that the two forms are not equally standardised. This difference in standardisation becomes obvious when we consider the mean scores obtained by the E and F-groups on the two forms of the test (see Table 7). From these mean scores we can derive mean gains and mean standardisation differences (shown in the last two columns of Table 7). It will be seen that the discrepancy in standardisation increases with reading level.

TABLE 7

MEAN GAINS AND STANDARDISATION DIFFERENCES DERIVED FROM MEAN SCORES OF E AND F-GROUPS ON RCA AND RCB.

School	Group	Mean score (months)		Mean gain (months)	Standardisation difference A-B (mths.)
		RCA	RCB		
P	E	98.4	97.1	4.7	6.0
	F	106.0	95.3		
Q	E	104.4	108.2	11.2	7.4
	F	122.7	104.2		
R	E	114.6	107.8	2.5	14.3
	F	123.2	111.4		
S	E	87.4	89.9	5.0	2.7
	F	99.8	92.1		

A method for dealing with this difficulty is described in Section B of the Appendix. By means of this method, we derived measures of the gains made by pupils at reading levels of 7.5 and 9.5 years. (It is to be remembered that in this case only about 1 month is to be deducted for maturation.)

It will be observed that at the 7.5 level, the gain in going from the one form to the other is only about half of the gain involved in being re-tested on the same form of the test.

With the data which we have presented in this and the last section, it is not very difficult to show that the gains made by the control pupils in our earlier remedial education experiment can be attributed entirely to maturation and practice effect.\* There is, therefore, no need for the *ad hoc* hypotheses favoured by Birch and the others.

TABLE 8

DIFFERENTIAL GAINS MADE BY E AND F-GROUPS ON TWO FORMS OF THE READING COMPREHENSION TEST.

School	Gain (months)	
	9.5 year level	7.5 year level
P	6.6	2.5
Q	11.4	6.8
R	3.4	4.6
S	7.5	2.8
Total . . . . .	7.2	4.2

In our earlier study, we included statistical regression as one of the factors contributing to the gain of our retarded pupils. The gain due to regression is not just a function of the reliability of the initial and final tests—as Birch would have it—but of the correlation between the initial and final test scores. This correlation is affected not only by the errors of measurement (with which the reliability coefficients are concerned), but also by real differential changes in the performance of the children in the inter-test period (and by differences in test content where the initial and final tests are not the same). In the present investigation, the same tests were used initially and finally and the inter-test periods were comparatively short. Regression effects were, therefore, found to be slight. On the whole, the same would apply to our earlier study.

#### IV.—COACHING.

A control group design enables the investigator to allow for such irrelevant factors as practice effect and regression, but in a remedial education experiment the *net* gains of the remedial group over the controls can still be inflated as a result of coaching. In our earlier study, we tried to demonstrate the coaching effect by showing that the 'true' (or net) gain for reading (as measured by the Schonell RC test) 'disappeared' when we applied an unstandardised test of different form and content.

Pringle and Gulliford threw doubts on this result and the conclusion we drew from it by criticising the validity of the unstandardised test we employed. But further evidence of a coaching effect was obtained when we re-tested both the remedial and control pupils a year after the finish of our experiment. The results are shown in Table 9. (The Schonell Graded Word test was included since our omission of this test from our experiment had been criticised.)

\* There is one exception. The large gain in going from Form A to Form B of the RC test, which we recorded in our earlier investigation, was spuriously high because of a difference in the standardisations which we employed for the two forms of the test. Instead of the gain of 7-10 months, which we recorded, we should have obtained something of the order of 4 months (as in the present investigation).



TABLE 9

MEANS FOR REMEDIAL AND CONTROL GROUPS (12 MONTHS AFTER EXPERIMENT IN REMEDIAL EDUCATION).

Test	Remedial Group (test-selected)		Control Group		Remedial Group (teacher-selected)		Control Group	
	yrs.	months	yrs.	mths.	yrs.	mths.	yrs.	mths.
RCB	9	10.3	9	9.5	9	2.1	9	4.3
GW	9	8.9	10	1.0	9	6.2	9	9.4

The differences between the remedial and control groups are quite non-significant statistically. The remedial pupils, therefore, failed to maintain the advantage in reading which they had shown at the end of the experiment.\* The transitoriness of the gain of our remedial group we would regard as evidence of coaching.

It will be seen that in our experiment and its follow-up we have used two criteria to detect the presence of coaching.

(a) Coaching is present when gains vary considerably with the type of test used to measure the basic skill to be acquired. Such variation in gains we would regard as evidence more of a variation in the acquisition of test-skills than of a genuine improvement in the basic skill.

(b) A gain can be attributed to coaching when it is transitory, i.e., when it disappears in a fairly short period of time—12 months or less.

Our second criterion represents a broadening in our conception of coaching as expressed in our original article and conforms more to the generally accepted idea of coaching.

Substantial confirmation of our results has been obtained by Dr. J. E. Collins, a former research fellow of this department.<sup>7</sup> Collins not only repeated our experiment on a larger scale, but also carried out a follow-up of twenty-four cases who had been treated at the Remedial Education Centre. The results of his experiment were identical with our own. Gains varied with the type of test used and were quite transitory, disappearing in three to twelve months. His follow-up of twenty-four past Centre cases also revealed a complete impermanence in the gains achieved by the Centre. The data which he obtained for these pupils are shown in Table 10.

TABLE 10

TWENTY-FOUR PAST CENTRE CASES—INITIAL, FINAL AND FOLLOW-UP TEST RESULTS.

Item	Data on entering Centre		Data on leaving Centre		Follow-up data	
	yrs.	mths.	yrs.	mths.	yrs.	mths.
C.A. ....	9	9	10	10	12	3
M.A. ....	10	7	11	10	13	3
Mechanical Reading .....	7	9	9	8	10	6
Reading Comp. ....	8	1	9	10	10	5
Spelling .....	7	0	8	3	9	0
Mechanical Arithmetic .....	8	6	9	8	10	9

\* The same result was obtained by W. K. Gardner, one of our Diploma students, who, at the time we were concluding our own experiment, re-tested the remedial pupils of the previous session. It is interesting to note that since Gardner did not choose his control pupils until a year after remedial instruction had taken place (on the basis of the data of the 'school surveys' which the Diploma students carried out each year), there was no question of the class teachers paying special attention to the controls (as Birch had suggested).

It would appear, therefore, that remedial education, at least as practised at the moment, has no permanent results. The gains which have been claimed for it are a product largely of practice effect and coaching. In other words, it accomplishes little more than is achieved by ordinary classroom teaching. To quote Collins, it represents an overlapping situation without essential change.

## VI.—APPENDIX.

- (a) *Method for estimating differential practice effects when pupils are re-tested on the same test.*

The methods used by Peel<sup>5</sup> and by Wiseman and Wrigley<sup>6</sup> are suitable only with large samples of pupils. By combining our school groups we might have obtained a large enough sample, but a necessary condition for this combination is that a covariance analysis of final and initial scores should show no difference between schools and this did not hold in our own case. We were, therefore, obliged to obtain differential practice effects for each school separately before averages for the four schools could be found.

Since our school groups were small, there was no possibility of establishing any non-linearity in the differential practice effects (as happened in Peel's investigation), but this did not matter since non-linearity was likely to be slight and to occur mainly at the higher levels (in which we were not interested).

Our method was simply to derive the equation of the 'equivalence line' for the final and initial scores, that is to say, the line showing score ( $\bar{Y}$ ) in the final test as a function of score ( $\bar{X}$ ) in the initial test. The equation of this line is

$$Y - \bar{Y} = \sigma_y / \sigma_x (X - \bar{X})$$

where  $\bar{X}$ ,  $\bar{Y}$ ,  $\sigma_x$ ,  $\sigma_y$  are the means and standard deviations of the initial and final scores, respectively  $k$ , for the school group.

This equation may be written in the form :

$$Y = b_0 X + c_0$$

where  $b_0 = \sigma_y / \sigma_x$  and  $c_0 = \bar{Y} - \sigma_y / \sigma_x \bar{X}$

Then, if there was no practice effect, the relation between  $Y$  and  $X$  would be simply  $Y = X$ . It follows that the differential practice effect  $P$  for any  $X$  is given by the equation

$$P = (b_0 - 1)X + c_0$$

In our investigation,  $P$  was found for  $X = 114$  and  $90$  months. It is these values which appear in Table 6.

- (b) *Method for estimating differential practice effects for two tests which are unequally standardised.*

In this case, it is necessary to have two groups in each school, one group receiving test (A) first and test (B) second and the other group receiving the tests in the reverse order. In our investigation, we denoted these groups by the letters E and F.

In the method we adopted, it is necessary to assume that while the two tests are unequally standardised there is a linear or approximately linear relation between equivalent ages as obtained by a school group on the two tests. This line we shall call the 'true' equivalence line. We must also, as in (A), assume that the differential practice effects are linearly related to level of ability.

Our procedure was then as follows :

- (i) The equation of the equivalence line for the (A) and (B) scores (in months) was obtained for both the E and F groups in each school. Then, if  $X$  and  $Y$  denote the (A) and (B) scores, the equivalence lines could be written in the form :

$$Y = b_1 X + c_1 \quad \text{(Group E)—(1)}$$

$$Y = b_2 X + c_2 \quad \text{(Group F)—(2)}$$

(ii) The 'true' equivalence line lies somewhere between (1) and (2). If we make the further assumption that the amount of practice effect (in months) in going from (A) to (B) is the same as that in going from (B) to (A), it is easy to show that the equation of the 'true' equivalence line is:

$$Y = \frac{b_2(1+b_1)}{1+b_2}X + \frac{b_2c_1+c_2}{1+b_2} \quad (3)$$

(iii) By subtracting (3) from (1), we obtain the differential practice effect P for any X.

$$P = \frac{b_1-b_2}{1+b_2}X + \frac{c_1-c_2}{1+b_2} \quad (4)$$

Using (4) we found the differential practice effects (differential gains) shown in Table 8.

It will be noted that, in all cases, differential practice effects were obtained relative to scores on Form A. Considerable divergencies between the published norms for (a) and (b) have long been apparent to users of these tests. Re-standardisations are lengthy and expensive, but by the application of equivalence-lines to the present data, we have obtained a re-standardisation for which we claim a high degree of accuracy. Briefly, the method consisted of (i) the use of the 'true equivalence line' for tests RCA/RCB (described above) to convert any score on one of these tests into the equivalent score—representing the same level of reading ability—on the other test; (ii) the application of the same method to the scores of groups ABCD to obtain the relation between GW and RCB; (iii) combination of the two conversion tables thus obtained, to give a table from which new norms for any two of the tests can be compiled on the assumption that the existing norms for the remaining test are valid. Since the standardisation of the graded word test is very generally accepted as sound, this was taken as a basis, and new norms were read off for RCA and RCB. The former were found to agree closely with the published norms, but Form B showed considerable discrepancies. The new norms for Form B which we derived by this method are shown in Table II. These are for both sexes together, but if we wish to separate the norms for boys and girls (as Schonell does for RCA and RCB, but not—oddly enough—for GW) this can be done by adding or subtracting the small adjustments shown.

TABLE II  
REVISED NORMS FOR FORM B OF SCHONELL'S READING COMPREHENSION TEST.

Score	Reading Age Yrs. Mths.	Score	Reading Age Yrs. Mths.	Score	Reading Age Yrs. Mths.	Score	Reading Age Yrs. Mths.
1 or 2	No score	12	9 1	22	11 3	32	13 0
3	6 11	13	9 4	23	11 5	33	13 2
4	7 2	14	9 7	24	11 7	34	13 4
5	7 5	15	9 9	25	11 10	35	13 6
6	7 8	16	10 0	26	12 0	36	13 8
7	8 11	17	10 2	27	12 2	37	13 9
8	8 2	18	10 5	28	12 4	38	13 11
9	8 5	19	10 7	29	12 6	39	14 0
10	8 8	20	10 10	30	12 8	40	14 1
11	9 10	21	11 0	31	12 10	41	14 2

*Sex Adjustment*

For scores 3 to 7 and 24 to 41 (Boys) add one month to Reading Age; (Girls) subtract one month from Reading Age.

For scores 8 to 23, inclusive (Boys) add two months to Reading Age; (Girls) subtract two months from Reading Age.



## VII.—REFERENCES.

- <sup>1</sup> CURR, W., and GOURLAY, N. (1953). An experimental evaluation of remedial education *Brit. J. Educ. Psychol.*, 23, 45-55.
- <sup>2</sup> BIRCH, L. B. (1953). Comments on 'An experimental evaluation of remedial education,' by Curr and Gourlay. *Brit. J. Educ. Psychol.*, 23, 56-57.
- <sup>3</sup> KELLMER PRINGLE, M. L., and GULLIFORD, R. (1953). A note on 'An evaluation of remedial education.' *Brit. J. Educ. Psychol.*, 23, 196-199.
- <sup>4</sup> CURR, W., and GOURLAY, N. (1956). Differences between testers in Terman-Merrill testings. *Brit. J. Stat. Psychol.*, 9, 75-81.
- <sup>5</sup> PEEL, E. A. (1952). Practice effects between three consecutive tests of intelligence. *Brit. J. Educ. Psychol.*, 22, 196-199.
- <sup>6</sup> WISEMAN, S., and WRIGLEY, J. (1953). The comparative effects of coaching and practice on the results of verbal intelligence tests. *Brit. J. Psychol.*, 44, 83-94.
- <sup>7</sup> COLLINS, J. E. (1956). *The Effects of Remedial Education*. Unpublished Ph.D. Thesis, Birmingham University.

(Manuscript received, September 28th, 1959.)

### III—PART-TIME CLASSES FOR YOUNG BACKWARD READERS.

BY H. H. HILLMAN AND R. L. SNOWDON

*(Handicapped Pupils' Section, Education Department, County Council of Durham.)*

SUMMARY. 1.—A scheme for the remedial teaching of reading was established to determine whether part-time teachers working with small groups of backward children could significantly improve the reading achievement of their pupils.

2.—130 children of junior school age had been tested and taught under this scheme by April, 1958. The remedial teachers adopted a predominantly 'phonic' approach.

3.—Comparison was made between the progress of backward children who attended remedial classes and children similarly handicapped who did not. Comparison was also made between the progress of children before and after a period of remedial teaching. The difference in both comparisons was found to be significant at the 0.01 level.

4.—The scheme was found to be acceptable to the schools from which the children came.

5.—The results appear to be roughly comparable with the results of similar (and more elaborate) remedial schemes and with the results of teaching retarded children.

#### I.—INTRODUCTION.

REMEDIAL classes for the teaching of reading to junior school children were started in Durham County during 1957 by the Education Committee. The purpose of this scheme was to relieve child guidance clinics of work which was purely educational, so as to leave them free to deal with children really needing specialist psychological help. The underlying assumption was that the great majority of backward readers in our junior schools are suffering primarily from an educational handicap and that emotional disturbance and social maladjustment in such pupils are due in almost every case to failure in the classroom.

There were four adjustment classes (as they were called) at each of two centres to deal with backward readers from nearby schools. Each class consisted of twelve children and met twice a week. A total of 130 pupils from twenty-six schools were dealt with during the period described in this article. Teaching was largely by qualified part-time teachers, principally married women, with infant or junior school experience, but otherwise with no special training in remedial methods. This was in accordance with the hypothesis upon which the classes were based. One full-time teacher was also employed under the scheme and was responsible for the selection of children and deciding when attendance should stop.

Children between 8 and 11 years old were admitted to an adjustment class if they were more than two years below the average attainment appropriate to their age, provided that they had not previously been ascertained as educationally sub-normal. It was not impossible for a child who was educationally sub-normal to be admitted because ascertainment had not been made before the date of entry. Nor did these classes attempt to deal with the bright child who, although reading to the average standard of an age-group, might be deemed to be retarded far below the attainment level of which he or she was capable. The 130 junior school children were not exclusively children with attainment far

below I.Q., nor exclusive of many children below average in intelligence. What we offered these children was more individual teaching than is possible in most primary schools to-day.

## II.—TESTING AND TEACHING IN PART-TIME CLASSES.

Selection for admission was made after each child had been fully discussed with the head teacher and class teacher concerned. No intelligence test results were available (except in the case of eight children referred by educational psychologists of the School Health Service), but Burt's Graded Vocabulary Test of Reading Accuracy<sup>1</sup> was administered to each pupil for the purposes of selection and the measuring of progress. This standardised test was chosen because it was considered that in the final analysis reading achievement springs from independence in the recognition of unknown words which itself depends upon developing phonic ability. The Burt test, though far from ideal, appeared to be more closely related to the growth of phonic knowledge than any other test on the market at the end of 1956, when the preliminary tests were given. The average amount of backwardness in reading, on entry to adjustment classes, as measured by the Burt test, was three years two-and-a-half months. Similarly, a pupil's readiness to leave was decided jointly by the adjustment class teacher and the child's own teacher and head teacher. In reaching this decision, the full-time teacher in charge of adjustment classes played the part of co-ordinator, visiting the schools and providing a Reading Age obtained with the Burt test.

Each remedial session was from two to two-and-a-half hours long and quite obviously children could not be expected to devote their whole time to reading only. The only request made of the teachers was that class activities should invariably be related directly to the reading improvement of each child. It was hoped that within the scope of this request each teacher would be free to develop her own ideas and orientate her instruction to the age and interest level and particular reading difficulties of the individual child. The basic reading series provided for each centre were Emily Smith's *Big Ben Readers* and Daniels and Diack's *Royal Road Readers*. In addition, there were a large number of supplementary reading books of which Gertrude Keir's *Adventures in Reading* series proved very useful. None of the seven teachers at work during the experimental period used methods which differed markedly from ordinary classroom procedures. All, however, adopted a consistently 'phonic' rather than a 'whole-word' or 'sentence-method' approach. It was adjustment class policy to respect the independence of the teachers' professional opinions and this agreement on method between organiser and teacher was fortuitous, not contrived. A wealth of materials (plasticene, coloured pencils, scissors, card, and so on) for the construction of reading apparatus and for related activities was available at each centre. This was used extensively by the teachers to enlarge and reinforce the reading and writing experience of their pupils and to vary their own rather formal 'phonic' techniques.

## III.—RESULTS.

The success of this type of remedial teaching as a means of overcoming backwardness in reading was measured primarily by the results of the Burt test (Table 1).

An average improvement in reading ability of eleven months for children who attended an adjustment class for one term, to two years five months for children who attended for three terms, would appear to be a satisfactory outcome of remedial teaching bearing in mind the following qualifying factors. Five children who were included for testing in the groups represented in Table 1



TABLE 1

IMPROVEMENT IN READING ABILITY OF CHILDREN ATTENDING CLASSES.

Improvement after Attendance FOR	No. of Children Tested	Least Improvement	Greatest Improvement	Average Improvement
One term . . . . .	118	Nil	3 yrs. 6 mths.	11 mths.
Two terms . . . . .	92	1 mth.	3 yrs. 2½ mths.	1 yr. 7 mths.
Three terms . . . .	29	8 mths.	3 yrs. 1 mth.	2 yrs. 5 mths.

were regarded as being unsuitable for this type of instruction. Of one of these children his head teacher remarked: " (He is) emotionally unstable and with a speech handicap is making slow progress in all subjects. I think he ought to be interviewed by a psychologist with a view to special school training." It was recommended to head teachers that these children should be reported to the School Health Service so that their disabilities might receive further investigation. So far as this affected the results shown in the above Table 1, entries in the column headed 'Least Improvement' represent the achievement of one or other of these five children in each case. The results obtained from these children were, of course, included in the populations from which the averages of improvement were calculated. The test results from which these averages were obtained were all provided by children with a record of over 60 per cent. attendance at an adjustment class. Their average attendance was 84 per cent.

It is not, however, sufficient to show that the children's reading made some improvement in order to justify this special educational provision, since it could be claimed that pupils would have progressed equally well if left in their own schools. In an attempt to meet this point, the following two studies were made:

(i) In Table 2 are shown the results of a comparison between the average improvements in reading ability of two groups of children matched on criteria of similarity of age, school, socio-economic background and degree of backwardness. Group A was made up of twenty-five children who spent at least two terms (Spring and Summer Terms, 1957) in ordinary junior schools, and in some cases considerably longer, without attending an adjustment class. The twenty-five children of Group B also spent two terms in ordinary junior schools but, in

TABLE 2

COMPARISON WITH CONTROL GROUP.

MATCHED GROUPS (25 children)	Average Age	Average Retardation	Average Improvement after two terms
GROUP A (In own Schools)	9 yrs. 6 mths.	3 years	4 months
GROUP B (In Adjustment Class) . . . . .	9 yrs. 7 mths.	3 years	1 yr. 11 mths.

addition, they travelled twice a week throughout this period to an adjustment centre for part-time classes. The children in both groups were tested with the Burt Vocabulary Reading Test at the beginning and end of the specified period. A test of the significance of the difference between the means of these two groups indicated that this result ( $S.E._a=0.1219$ ) is significant above the 0.01 level.

(ii) As a further check on this finding, however, a different type of control group was used. In this study the average improvement in reading ability of twenty-one children who spent two terms in their own schools without adjustment class help was found to be five months, compared with an average improvement in reading age for the same children, after attending adjustment classes for two subsequent terms, of one year ten months. These two means were tested for significance and were found to be such ( $S.E._a=0.154$ ) that it is most unlikely the difference between them occurred by chance.

So far attention has been directed to a statistical analysis of the achievements of adjustment classes. In practice, however, the success of the centres was measured by many factors which cannot be represented numerically. In particular, the interest and active co-operation of the children themselves, their parents, head teachers and class teachers were evidence of the value of this type of remedial teaching. It is insufficient to provide a specialised educational service which is validated by statistical proofs, unless the head teachers who are free to avail themselves of this service are of the opinion that it performs a necessary function and that the means chosen are both practical and efficient. In particular, the willingness of parents to allow their children to attend adjustment classes depended entirely on the good relations established by the head teachers of the children's own schools. Considerable importance may, therefore, be attached to the fact that of the twenty-seven head teachers concerned, all expressed favourable opinions of the work of the classes and none contrary.

Various education authorities, as well as individual investigators, have undertaken researches into the value of remedial class teaching. Despite this, it was difficult to find reports of this character which were directly comparable with the present report. For interest, however, several comparisons were made with work by other teachers and it is contended that the rate of improvement in Durham compared very favourably with rates obtained elsewhere, e.g., Manchester,<sup>2</sup> Schonell,<sup>3</sup> Birch,<sup>4</sup> and Miss Friedmann's retarded children.<sup>5</sup>

#### IV.—CONCLUSION.

The relevance of these results to those administrators and teachers concerned with backward readers in our junior schools appears to us to be two-fold. First: we conclude that remedial classes, economically organised as described above and staffed by part-time teachers without specialist training, were able, in Durham County, significantly to reduce the reading backwardness of the majority of children referred to them. Although it cannot be stated that better results might not have been obtained by a more elaborate provision, it would seem that Durham results bear comparison with published reports of experiments both similar and more complex. Second: we conclude that there is some indication that a more individual approach to ordinary backward readers of junior school age can produce educational results apparently as good as a similar approach to a selected group of retarded children.

ACKNOWLEDGEMENT.—The authors wish to express their gratitude to Mr. G. H. Metcalfe, Director of Education, for the County Council of Durham, for his kind permission to publish these results.

## V.—REFERENCES.

- <sup>1</sup> BURT, SIR CYRIL (1949). *Mental and Scholastic Tests*. London: Staples Press. 367-369.  
<sup>2</sup> CITY OF MANCHESTER EDUCATION COMMITTEE (1957). *Annual Report on the School Health Service*. Manchester: Co-operative Press, Ltd., 42-46.  
<sup>3</sup> SCHONELL, F. J. (1949). The development of educational research in Great Britain: Part V—Handicapped children. *Brit. J. Educ. Psychol.*, 19, 82-97.  
<sup>4</sup> BIRCH, L. B. (1949). *Educ. Rev.*, 1, 107. Cited by VERNON, M. D. (1957). *Backwardness in Reading*. Cambridge University Press, 179-180.  
<sup>5</sup> FRIEDMAN, S. (1958). A report on progress in an L.E.A. remedial reading class. *Brit. J. Educ. Psychol.*, 28, 258-261.

(Manuscript received, June, 19th, 1959.)



## IV—THE EFFECTS OF REMEDIAL EDUCATION ON YOUNG CHILDREN'S READING ABILITY AND ATTITUDE TO READING.

By J. DUNHAM

*(Summary of thesis approved for the degree of M.Ed. in the University of Manchester, May, 1959.)*

### AIMS.

The main aim of this investigation was to obtain experimental evidence about the effectiveness of remedial education. Evidence of this kind was thought to be particularly valuable because special provision for remedial education and criticism of remedial education have both been increasing recently in this country. Two criteria of effectiveness were used :

- (a) Improvement in reading ability.
- (b) Improvement in attitude to reading.

The question which the main part of the investigation sought to answer was : What are the effects of remedial education on young children's reading ability and attitude to reading ?

The thesis also contains some findings on the following controversial topics .

- (i) The progress of retarded children not given remedial education.
- (ii) The selection of children for remedial education.
- (iii) The use of non-verbal tests of intelligence in selection.

### METHODS.

The remedial group (A) consisting of twenty severely retarded 9-year-old children of about average intelligence, took part in remedial reading lessons once or twice weekly for six months, amounting to about twenty-two hours of remedial work. They were paired, child for child, with a control group of children on the waiting list for remedial education, according to age, intelligence, reading ability and the socio-economic area of the school. The remedial children worked in small groups of not more than six as part of the normal work of an L.E.A. Remedial Education Service.

Reading ability was measured on the Burt (re-arranged) Word Recognition Test. Retardation was measured by the reading test, the N.F.E.R. Non-Verbal Test 5 and the Verbal Scale of the W.I.S.C. (U.K. version). Attitude to reading was measured on a Thurstone-type attitude scale, prepared by the writer.

The 'real' gain in reading of the experimental group was the difference between the mean gains of the experimental and control groups. This was the increment credited to remedial education. Similarly, the 'real' improvement in the attitude to reading of the experimental group was the difference between the means of the improvement in attitude to reading of the experimental and control groups.

In order to investigate the prediction of success in remedial education, because it is an important aspect of the problem of the selection of children for remedial work, the reading gains of the children of the experimental group were correlated, using Spearman's rank order formula, with their R.A., C.A., I.Q., and Attitude Score at the beginning of the investigation and with their changes in attitude during the investigation.

To compare the effect of using a verbal with that of a non-verbal test of intelligence, in the selection of children for remedial education, the respective results of the experimental group, when tested on the N.F.E.R. Non-Verbal Test 5 and the W.I.S.C. (V) were studied.

To assess changes in attitude to reading, an attitude scale constructed by the writer in an earlier research was used after some modification and improvement. This scale was of the Thurstone type ; care was taken in its construction to ensure that the more ambiguous items were excluded, and that the remaining items were fairly evenly spaced by scale value. Three practice statements were added to assist the children to understand the procedure. The scale values of the twenty items finally selected ranged from 0.4 to 10.4, with high SV representing a favourable

attitude to reading, and the attitude score was the mean scale value of the statements with which the child agreed. The test-re-test coefficient of reliability of the scale was found to be  $+ .77$ . The validity coefficient of the scale obtained by a comparison with a class teacher's rank order was  $+ .59$ . The scale was administered individually and orally.

# RESULTS.

## 1.—Improvement in Reading Ability.

Group	No. of children	At beginning			At end		
		Mean C.A. (y : m.)	Mean I.Q.	Mean R.A. (y : m.)	Mean R.A. (y : m.)	Mean Gain (months)	"Real" Gain (months)
A	20	9 : 5	99	5 : 8.9	6 : 7.9	11.0	6.4
B	20	9 : 5	99	6 : 5.9	6 : 10.5	4.6	—

The 'real' gains were significant at the 1 per cent. level.

## 2.—Improvement in Attitude to Reading.

Group	Mean Attitude Scores		Mean Improvement in Attitude	"Real" Improvement in Attitude
	At beginning	at end		
A	+6.8	+7.6	+0.8	+0.5
B	+7.3	+7.6	+0.3	—

The 'real' improvement in Attitude to reading was not significant.

## 3.—The Selection of Children for remedial education.

The following rank order correlation with the reading gains of Group A were obtained :

C.A. ....	-0.07	Extent of retardation .....	-0.11
I.Q. (N.F.E.R. 5) .....	+0.21	Attitude Score .....	+0.10
R.A. ....	+0.16	Changes in Attitudes .....	+0.27

None of the correlations was significant.

## 4.—Comparison of the test results of Group A when tested on the N.F.E.R. Non-Verbal Test 5 and the W.I.S.C. (V).

Test	Mean I.Q.	Mean M.A.	Rho.	Mean Retardation	Range of Retardation	Rho. with Reading gains
NFER 5 ..	99	9 : 4	+0.69	3 : 6	3 : 10	+0.21
WISC (V) ..	94	8 : 11	—	3 : 2	3 : 0	+0.22

Mean I.Q. when tested on the non-verbal test was significantly higher (at 5 per cent. level) than mean I.Q. when tested on the verbal test.

## SUMMARY OF CONCLUSIONS.

1.—There was evidence that remedial education, as organised in this investigation, was, on the whole, effective in improving the reading ability of severely retarded 9-year-old children of about average intelligence.

2.—There was evidence that the favourable effect of remedial education, as organised in this investigation, on the attitudes to reading of the same children was much less than the effect on reading ability. If this finding is reported from other investigations it would appear to suggest that more research ought to be directed to the problem of getting the retarded reader to approach the reading situation in a wholesome way, particularly when his home and environmental conditions seem very unfavourable.

3.—The finding that in the control group some children made very little progress in reading and some even lost ground, suggests their urgent need of remedial education. But, in fact, that some reasonable gains were also recorded suggests the need for research to attempt to formulate satisfactory selection procedures so that retarded children who would probably make progress without it are not given remedial education.

4.—The need for further research into selection procedures is also suggested by the finding that the reading gains of the experimental group were associated at least as closely with improvements in attitude to reading as with I.Q.

5.—There was evidence that, on the whole, the test scores of the children of the remedial group were significantly higher, when tested on the N.F.E.R. Non-Verbal Test 5 than on the Verbal Scale of the W.I.S.C. This suggests that their general level of verbal ability was on the whole, *lower* than their general mental ability level. Because of the reported close relationship between reading and verbal or language development, particularly in the early stages of learning to read, this finding appears to support the suggestion that retardation should be diagnosed as a discrepancy between the general level of verbal development (as shown by a verbal or 'language' I.Q. or Mental Age) and the slower growth of reading skills. For this the W.I.S.C. (V) would appear to be a more appropriate test than the N.F.E.R. Non-Verbal Test 5. On the other hand, there has been no evidence that the W.I.S.C. (V) would be more useful in predicting success in remedial education than the N.F.E.R. Non-Verbal Test 5.

6.—Evidence is presented that a Thurstone-type attitude scale, when administered individually and orally, can be successfully used to measure the attitude to reading of 9-year-old pupils retarded in reading. As attitude scales have been used almost entirely with secondary school pupils, this may be a useful finding.

# V—A STUDY OF MOTIVATION IN REMEDIAL READING

By GEOFFREY R. ROBERTS

(Summary of M.A. thesis, University of Wales, 1958.)

## 1.—INTRODUCTION.

Three forms of motivation, each combined with a suitable method of teaching children to read, were compared under normal classroom conditions. The experiment, which lasted 6 months, was carried out twice. On the first occasion, eighteen children took part, divided into three groups of six; and on the second occasion thirty-three children took part, also divided evenly. However, nine of these thirty-three had taken part in the experiment once, so their results at the second attempt were excluded. Thus, for statistical calculations, the numbers were twelve on Method A, sixteen on Method B and fourteen on Method C. Twenty weeks, with five half-hour lessons per week, were devoted to work on the methods in both periods.

All the children were from low cultural backgrounds. They were placed in roughly balanced groups according to their general reading ability.

Initial Means:

Group A—CA. 10y. 2m., Range 9y. 8m. to 10y. 9m.  
I.Q. 98, Range 70 to 114, SD.=13.7.

RA. 8 y 6m., Range 6y. to 11y.

Group B—CA. 10y. 2m., Range 9y. 4m. to 11y.  
I.Q. 100, Range 80 to 127, SD.=13.2.

RA. 8y. 5m., Range 6y. to 11y. 2m.

Group C—CA. 10y. 2m., Range 9y. 4m. to 10y. 11m.  
I.Q. 104, Range 89 to 122, SD.=9.8.

RA. 9y. 2m., Range 7y. 2m. to 12y. 2m.

## 2.—DESCRIPTION OF THE THREE METHODS.

METHOD A: The basis of motivation was curiosity and creative drive, and the method devised by Grace M. Fernald was used, with a few minor alterations. The first step was to restore the children's confidence in their ability to learn (reconditioning). The power of this method was demonstrated by getting any child to choose a difficult word. It was written on a piece of paper and the child was told to trace it with his fore-finger on the desk, pronouncing its component parts simultaneously, e.g., im-port-ant. Individual letters were not sounded. The tracing and vocalization continued until the child could write the word without copying; then all the children were invited to try their skill in learning a word.

The main task of writing stories and reading them now began. The child chose a topic and wrote his story, using one side of each piece of paper. Any word which he could not spell was written on the back of his paper, in fairly large letters, and he learnt to spell and recognise it by the method of tracing—vocalization—writing. The child entered all new words in a personal word book and occasionally was tested on them. When the story was finished, it was reprinted by the teacher, within twenty-four hours, pasted in the child's personal story book, and read to the teacher. No restriction was placed on the words used and gradually the tracing diminished. Any words mis-spelt in the story were learned after reading the reprinted story.

This process continued throughout the experiment and the children were allowed to read their stories to other children and also to read other children's stories. This was very popular. The children illustrated their stories. If anyone wished to read he could choose any available book, and new words were learnt as above.

METHOD B: Motivation depended upon gregariousness. Co-operative effort was frequently possible. Competition was eliminated and the work of the children was never compared. The method used was based upon those advocated by A. E. Gates and F. J. Schonell for general use. It consisted of (1) a word enrichment programme; (2) co-operative reading; (3) individual reading; and (4) incidental phonic analysis.



The word enrichment programme was planned and occupied a third of the time. It took the form of non-competitive games and puzzles. These included: (1) finding the missing letter or word; (2) finding the word where the letters have been jumbled up; (3) selecting words relating to a particular topic; (4) re-writing sentences where the words have been jumbled; (5) learning a list of words; (6) crossword puzzles. Exercises were given which included: (1) writing words relating to a topic; (2) finding words beginning with a given letter (no dictionaries); (3) matching words and sentences to pictures; (4) matching words and sentences printed on cards against the same word or sentence in a bundle of cards or in a book. No formal lessons were given, but the essential points of the games and exercises were emphasized.

Co-operative reading included: (1) reading aloud in pairs or to the teacher; (2) a more advanced reader helping a less advanced reader; (3) reading in groups (soon discarded as futile).

The remainder of the time was allocated to individual silent reading. The children worked through a graded list of 146 books, beginning on an appropriate book. Some incidental phonic analysis of new words was employed. The sound and form of the word was pointed out, but the child was left to employ his own method of learning the word.

**METHOD C:** The basis for the motivation of Group C was competition. All the exercises and much of the individual reading were on a competitive basis and a system of awards by stars was used, with a prize at the end of each term.

An essentially phonetic method was used, based upon lessons designed to eliminate reading errors classified by Marion Monroe as: faulty vowels, faulty consonants, reversals of letters and sentences, additions and omissions of sounds and words, substitution of words, repetition, and refusal to attack a word. Monroe's basic approach was followed, though her method was not used in its entirety. A rough profile of errors of each child showed a general tendency towards greatest deficiencies under 'faulty vowels and consonants, and reversals,' and prominence was given to their correction.

Method C consisted of (1) Lessons and drills of deficient elements; (2) Competitive games and exercises connected with these lessons; (3) Individual reading of the graded readers used by Group B, with competitions in oral reading and word recognition. Formal lessons were given and great emphasis was placed on all sounds, especially of letters. First, the various vowel sounds were dealt with and then the diagraphs and 'y.' Next came the consonants and then common reversals. Blends of two and three letters, silent letters, and finally common word-endings, e.g., tion, were illustrated. Other deficiencies were corrected individually. Formal instruction in each lesson lasted for five to ten minutes. Competitive games and exercises followed and involved thinking of words, finding words; underlining specific sounds; placing words in sentences, and pronouncing words. Some games were not necessarily connected with the lessons and included crosswords, finding the missing letter or word, etc. Individual reading occupied a little under half the time spent on this experiment and stars were given for good and improved reading.

### 3.—RESULTS.

Two tests were used—Sleight's Non-Verbal Intelligence Test and Schonell's Graded Word Reading Test. Both tests were given at the beginning of the experiment and the reading test was given again after ten weeks and finally after twenty weeks. Testing was kept to a minimum.

The mean improvement in Reading Age for each group was: Group A=15.7 months, SD.=6.5; Group B=7.9 months., SD.=4.2; Group C=10.5 months, SD.=5.2.

The differences between the means, when tested by Sir Ronald Fisher's 'small sample' statistical technique, showed that the results of Group A were significantly better than those achieved by either Group B ( $t=3.66$ ,  $P=<0.2$  per cent) or Group C ( $t=2.17$ ,  $P=<5$  per cent.). The results of Group C were not significantly better than those of Group B ( $t=1.45$ ,  $P=\text{about } 20$  per cent.), although with slightly bigger numbers of children it appears probable that they would be.

The numbers making 6 months improvement or less were : Group A—1 out of 12 ; Group B—6 out of 16 ; Group C—4 out of 14. The numbers making more than 12 months improvement were 8, 2 and 4, respectively.

A comparison was made between the M.A. and R.A. of each child at the beginning and end of the experiment. The mean reduction in retardation in reading was : Group A—9.6 months ; Group B—1.8 months ; Group C—4.1 months. The differences again showed the results of Group A to be significantly superior to both Group B ( $t=3.63$ ,  $P<0.2$  per cent.) and Group C ( $t=2.25$ ,  $P<5$  per cent.), with Group C again not quite significantly better than Group B ( $t=1.30$ ,  $P$ =about 20 per cent.).

Thus, Method A was shown to be very much the best method, and Method B the worst.

Rank correlations between initial R.A. and the amount of improvement shown during the experiment showed Group A =  $-0.3$  ; Group B =  $+0.2$  ; Group C =  $+0.1$ . Although not one of these correlations by itself is significantly different from zero, the difference between the correlations coefficient for Group A and the other two coefficients shows a tendency towards significance. Thus, Method A tends to favour the poorest readers more than either B or C does.

In Method C it was the best readers at the start who showed the greatest success in the competition (correlation coefficient of initial R.A. and numbers of stars gained =  $+0.9$ , and success in the competition was not related to the amount of improvement (correlation coefficient  $+0.2$ ).

Those children who had formed Groups B and C during the second occasion of the experiment were transferred to Method A for six weeks. The mean improvement in R.A. for each group of eleven children during this short period was : Group B—5.9 months, range 0 to 12 months, S.D. =  $3.7$  ; Group C—5.6 months, range 0 to 19 months, S.D. =  $4.7$ .

#### 4.—DISCUSSION AND CONCLUSIONS.

The results show that Method A is undoubtedly superior to the other two methods as a means of teaching reading to retarded readers between the ages of 9 and 11 years. The means of implementing a method could not be separated easily from the underlying motivation, therefore, it is possible only to give impressions of their relative merits.

The advantages of Method A are : (1) It is concerned with the interests of the child. (2) It increases self-confidence and self-esteem through writing, reading, and displaying their work. (3) By actively engaging the child, it prevents reading from becoming passive and purely mechanical. (4) Reading and writing are combined naturally, which increases the fluency of both. The children read their own stories for the first time with greater fluency than they read other stories. (5) The preliminary reconditioning instilled a tremendous and sustained burst of confidence, because the children possessed an effective and congenial method of learning new words. Tedious word lists were avoided and new words were discovered in a natural context and could be learned in association with a vivid experience. The method of learning words was simple and, whilst conforming to a single pattern, it incorporated visual, auditory, tactile and kinaesthetic means. (6) The child is induced to concentrate upon the form of a new word sufficiently long enough for him to learn it.

Disadvantages of Method B. (1) Gregariousness did not necessarily lead to co-operation. Rarely were all the children in a group working to capacity. (2) Learning tended to become the prerogative of the least retarded children, who tried to dominate rather than help the others. (3) It is difficult to obtain a suitable range of books for the older retarded child. (4) Following a scheme of graded readers involved the child in reading some books of little or no interest to him. (5) The children were always keen to change from reading to the word enrichment programme, although it failed, compared with the other methods, because it was remote from the child's reading and the transfer of words learnt in this way was a slow process.

Disadvantages of Method C : (1) The competition failed in that it did not help those whose requirements were greatest, even though the competitions were 'arranged' so that they could win stars. Furthermore, the star chart was a

constant reminder to them of their failure. (2) Undue anxiety, shown by fidgetting, pencil biting, and facial contortions, was aroused in the competitions. Even so, the children were keen to do the competitive exercises, which suggests that the act of 'doing' was a stronger source of motivation than the competitive urge. (3) The children easily became bored during the formal lessons, because the approach to words was too technical, dealing with unintelligible segments and letter sounds, rather than the sound of syllables.

General conclusions: A major cause of retardation was an inability to learn new words. Method B was the least effective in rectifying this, because, as in Method C, the word enrichment programme was separated from the child's reading. Method B had the further disadvantage of allowing the children to attack new words in a haphazard manner. It left the child to find its own way of learning new words. Method C gave the children a set form of attack upon new words, but the presentation lacked the vivid meaningful associations of Method A. In Method C, words were treated in isolation and were not closely connected with something of immediate interest to the child. In Method A, new words were a small obstacle in a large field of interest, whereas in Method C words became a huge isolated obstacle.

Method A, in contrast to Method C, did much to dispel anxiety. Method B and C did not possess the propensity for 'ego-involvement' that Method A did. In Method A the motivation came from within the child, whereas in B and C it was externally contrived. The children of Group A were thinking and writing about things which interested them; they could see their personal story book growing; and they could choose their own books. But the children of Group B were too young to co-operate, and for Group C competition merely favoured the more able. Lastly, Method A is superior as an instrument of class instruction because it acts as a remedy to a wide range of deficiencies, it readily motivates the majority of children, and it holds their interest over a long period.

# OTHER SUMMARIES OF RESEARCHES REPORTED IN DEGREE THESES

## The Number Concepts of a Group of E.S.N. Children.

By J. B. MANNIX

(Abstract of thesis submitted for the degree of M.Ed. in the University of Manchester,  
September, 1959.)

### AIM.

The aim of the research was to discover whether E.S.N. children behave similarly to or differently from the normal children tested by Piaget in the construction of certain concepts of number and quantity.

### METHOD.

Eight of the tests described by Piaget in *The Child's Conception of Number* were given to forty-eight E.S.N. children. The children were selected to represent a range of mental age from 5 years to 9 years. Three boys and three girls represented each six-month period of mental age from 5.0 to 8.11.

The tests were chosen so as to sample those performed by Piaget. There was a test from each chapter of *The Child's Conception of Number*, except those concerned with Seriation, Ordination and Cardination, and Additive Composition of Classes. The tests selected are shown below in the order in which they were given. (Page numbers refer to *The Child's Conception of Number*.)

- 1.—Conservation of Discontinuous Quantities (with initial correspondence) p. 25
- 2.—Conservation of Continuous Quantities ..... p. 4
- 3.—Additive and Multiplicative Composition (Problem III, Measuring) p. 222
- 4.—Spontaneous Correspondence—Single Rows ..... p. 74
- 5.—Provoked Correspondence (Correspondence between Eggs and Egg Cups) ..... p. 49
- 6.—Co-ordination of Relations of Equivalence (Eggs and Cups) ..... p. 204
- 7.—Additive Composition of Number—Equating of Quantities ..... p. 190
- 8.—Additive Composition of Number—Relations between Parts and Whole ..... p. 187

In giving the tests, Piaget's instructions were followed closely. The tasks were standardised and the children were asked standard questions, but supplementary questions were often asked to elicit further information. The tests were given individually.

### RESULTS.

*Evidence of Development.*—In every test, responses were found which could be readily identified as 'Operational' (Stage 3) responses, according to Piaget's criteria. Likewise, 'Pre-Conceptual' (Stage 1) responses were identified in all the tests. At both these stages, the children's responses were very similar to those quoted by Piaget.

The evidence for an intermediate stage ('The Period of Intuitive Thought') was less convincing. Some kind of intermediate stage was found in five of the eight tests, though only three of these showed second-stage responses very similar to those reported by Piaget. In the two tests of Additive Composition, there was no evidence of an intermediate stage, and in the remaining test there was a small and doubtful group of second stage responses. Evidence of logical multiplication of relations at this stage, upon which Piaget bases much of his hypothesis, was only slight and occasional.



*Counting and Correspondence.*—The E.S.N. children used counting to a greater extent than the normal children tested by Piaget, even when the test situation strongly suggested correspondence. Correspondence was used when counting was made difficult. The use of either was typical of the more advanced children, though counting was the preferred method.

*Scaling of the Test Results.*—Since some of the test results could not be classified into Piaget's three stages of development, it was only possible to scale all the test results together on a basis of two categories—'Operational' responses (Stage 3) and 'Pre-Operational' responses (Stages 1 and 2). Guttman's technique of scalogram analysis was applied and the responses were found to fall into seven scale types. The scalogram showed little evidence of disruptive factors and a reproducibility co-efficient of .94 was obtained. Though this figure met Guttman's requirements for 'scalability,' it was known to be spuriously high because only two categories had been used. However, it was thought that near-scalability had been reached and that the following tentative conclusions might be drawn: First, the fact that a child is operational in one test does not imply that he is operational in others; second, when the mean mental ages for each scale type are considered, no very close agreement is found between progression in the test responses and progression in mental age:

	Scale Type	Mean M.A.
Completely Pre-Operational	1	5.8
	2	6.8
	3	7.9
	4	6.10
	5	7.2
	6	8.6
Completely Operational	7	7.11

The scalogram also showed that the mental age at which children passed into the stage of Concrete Operations varied widely from child to child. It was possible to say, however, that no child with a mental age of less than 6.8 was at the stage of Concrete Operations in all the tests; and that no child with a mental age of more than 6.5 was at the Pre-Operational Stage in all the tests.

*Relationship of Scaled Test Responses with Mental and Actual Age.*—Spearman's rank correlation was used to find the relationship between

- (a) Scaled Test Responses and Mental Age ( $P = +.61$ ).
- (b) Scaled Test Responses and Actual Age ( $P = +.52$ ).

## CONCLUSIONS.

- 1.—The behaviour of E.S.N. children in the construction of certain number concepts is, on the whole, similar to that of the normal children tested by Piaget.
- 2.—Test behaviour suggests that they pass through the three stages of development described by Piaget.
- 3.—The period of development from the first to the third stage corresponds to a span of mental age of 5 years to 8 years.
- 4.—The mental age at which subnormal children pass from one stage to the next varies widely from child to child.
- 5.—The development stage is not necessarily constant for the same child from test to test.
- 6.—The absence of evidence of an intermediate stage may mean that such a stage does not exist or may indicate a lack of discrimination in the test.
- 7.—Counting appears to be the method preferred by E.S.N. children when dealing with a number problem, but correspondence may be substituted if the test situation makes counting difficult.

**A Study of the Reliability and Validity of the Children's Apperception Test.**

By P. V. REDDY, M.A., DIP.REM.ED.

*(Abstract of Ph.D. Thesis, University of London, Institute of Education, 1959.)***INTRODUCTION.**

The Children's Apperception Test (C.A.T.) is a projective technique developed by Bellak and Bellak<sup>1</sup> to investigate the personality problems of children between 3 and 10 years of age. The test consists of ten cards with animal pictures. The basic assumption underlying the test is that the children, by making stories about the pictures, identify themselves with the characters and thus indirectly reveal their personalities. Research with this pictorial projective technique so far has been somewhat scanty. Up to the present moment no systematic attempt has been made to check the reliability of the test or to validate the interpretation of its responses. This investigation is aimed to study these attributes of the C.A.T.

**SUBJECTS.**

The subjects in this study consisted of one-hundred 6 and 7-year-old children selected at random from a junior mixed school. An equal number of each sex were selected in each age group. In addition, nineteen male and six female deprived subjects whose ages ranged from 6 to 7 years were selected from two L.C.C. Homes for deprived children.

**TEST MATERIALS USED.**

(1) *The Crichton Vocabulary Scale* (1950) was used to measure the intelligence level of the subjects. The distribution of the I.Q.s. of the three groups of subjects is given in the table below. The 6 and 7-year-olds are referred to as Group I and Group II, respectively, in the table.

TABLE I

DISTRIBUTION OF I.Q.s. OF NORMAL AND DEPRIVED SUBJECTS ON THE CRICHTON VOCABULARY SCALE (1950).

I.Q. Range	Group I		Group II		Deprived Group	
	Male	Female	Male	Female	Male	Female
116—120	2	2	1	1	2	1
111—115	2	1	4	1	2	0
106—110	3	6	4	3	1	1
101—105	9	2	4	5	3	0
96—100	5	5	6	6	6	2
91—95	4	9	6	9	5	2
N .....	25	25	25	25	19	6
Mean I.Q. ...	103	101.2	102.4	99.8	98.3	
Std. Deviation	7.20	8.25	7.80	7.05	7.40	

(2) *Teachers' Rating Scale.* A rating scale was designed to enable teachers to assess the personality characteristics of the pupils who were selected for this study. The data thus obtained was used as an external criterion in validating the interpretations of the C.A.T.

(3) *The Children's Apperception Test* was administered on the lines suggested by Bellak. The manifest content of the subject's responses to the test was analysed by means of a scoring scheme designed by the writer. Bellak's scheme was found to be unduly complicated and too psycho-analytically oriented.

(4) *Lydia Jackson's Test of Family Attitudes.*<sup>4</sup>

# METHODS OF STUDY AND RESULTS.

(1) *Reliability.* The reliability of the C.A.T. was approached from three aspects :

(a) *Test-Retest coefficient* of the C.A.T. was obtained by re-administering the test to fifty 7-year-old subjects after an interval of three months. The correlation coefficient of sixteen personality variables ranged from +0.11 to +0.81 with an average of +0.45.

(b) *Score-Rescore reliability.* To determine the re-score reliability of the C.A.T., records of the fifty 6-year-old subjects were re-scored by the writer six months later and the resulting coefficients ranged from +0.51 to +0.95 with an average of +0.82 for all the scoring categories.

(c) *Reliability by the Matching Technique.* Three judges were asked to match two sets of personality sketches drawn up from the test-retest C.A.T. records of eighteen male and eighteen female subjects. The seventy-two sketches were divided at random into six groups with six pairs in each group. A statistical analysis of the matchings was made and an average coefficient of +0.578 ± 0.078 by Vernon's formula<sup>2</sup> was obtained.

(2) *Validity.* Construct validity and concurrent validity procedures were applied.

(a) *Construct Validity.* (i) *Factorial Validity.* The two sets of intercorrelations of the C.A.T. personality variables were separately analysed by Thurstone's (simple summation) method and three bi-polar factors were extracted for each age group.

Group I factors : (i) Cheerfulness *versus* Depression ; (ii) Extroversion *versus* Introversion ; (iii) Self assertion *versus* Self-submission.

Group II factors : (i) Cheerfulness *versus* Depression ; (ii) Extraversion *versus* Introversion ; (iii) Adequacy *versus* Inadequacy.

(ii) *Comparative study of normal and deprived subjects.* The C.A.T. responses of the twenty-five deprived subjects were compared with those of a control group. The responses of the two groups were analysed and compared on the basis of the following variables : personality variables, inter-familial attitudes, anxieties and fears and outcome of stories. On the whole the differences between the scores of the two groups on the above variables were found to be significant. The sign test results of the significance of the differences between the scores of the personality variables are given in the table below. Personality variables with a probability value of .05 or less, should be taken as significant.

TABLE 2  
SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE PERSONALITY VARIABLE SCORES OF  
NORMAL AND DEPRIVED SUBJECTS (TWO-TAILED).

	Personality Variable	High Score	P
1	General Emotionality .....	D	.286
2	Extroversion .....	N	.004
3	Introversion .....	D	.210
4	Self Assertion .....	N	.022
5	Self Submission .....	N	.134
6	Acquisition .....	N	.030
7	Sociability .....	N	.002
8	Cheerfulness .....	N	.002
9	Depressive .....	D	.062
10	Anxiety .....	D	.002
11	Feeding .....	N	.504
12	Ego strength .....	N	.002
13	Toileting .....	N	.096
14	Suprego Strength .....	N	.168
15	Mother Identification .....	N	.204
16	Father Identification .....	D	.504

N = Normal Group with high scores.  
D = Deprived Group with high scores.

(iii) *Changes in Personality.* The C.A.T. was used to measure changes in personality of fifteen subjects who were considered by their teachers to have changed most in their personality characteristics over a period of one year. Two sets of C.A.T. records and two sets of teachers' ratings—one before and one after the change—were available for each of the fifteen subjects. The data were analysed and the changes were shown by means of profiles. In the majority of cases the changes in the C.A.T. graphs appear to correspond with those shown in teachers' ratings graphs.

(b) *Concurrent Validity.* (i) *Validation of C.A.T. against Teachers' Ratings.* The correlations between Teachers' Ratings and C.A.T. personality variables were determined and a factor analysis of the intercorrelations resulted in the extraction of four factors. On the basis of the factorial evidence it seems that the C.A.T. is useful in measuring such personality characteristics of 6 and 7-year-old subjects as cheerfulness, sociability, ego strength, assertiveness and emotional stability.

(ii) *Validation by the Matching Technique.* Personality sketches of ten subjects drawn from their C.A.T. records were divided into four sets with five sketches in each set and were given to two judges to match with their own impressions of the same subjects. The two judges respectively, obtained an average of 50 per cent. and 35 per cent. of correct matchings. This represents an average coefficient of  $+0.624 \pm .160$  by Burt's formula<sup>3</sup> and  $+0.570 \pm .076$  by Vernon's formula.<sup>2</sup>

(iii) *The C.A.T. and the Family Attitudes Test Compared (F.A.T.).* The responses of forty-one subjects to the C.A.T. and to Jackson's Test of Family Attitudes<sup>4</sup> were compared to determine whether any significant association between the variables of the two tests existed. The results on the whole indicate a moderate degree of agreement between the C.A.T. and the F.A.T. Discrepancies between the variables may be attributed to the differences between the two tests.

#### DISCUSSION.

In general, this investigation has proved the C.A.T. to be fairly reliable and useful technique. The reliability results on the whole are quite comparable to those reported by other investigators with projective techniques. The findings of this study suggest the need for a shorter test-retest interval and for the investigation of inter-tester reliability. Further research is needed to ascertain more accurately the relationship between the personality characteristics of subjects and their responses to the C.A.T. An investigation into such personality problems as orality, sibling rivalry, toileting problems, etc., which the C.A.T. is specifically designed to study (and which were not available in the teachers' ratings) is suggested.

#### BIBLIOGRAPHY.

- <sup>1</sup> BELLAK, L. (1954). *The T.A.T. and the C.A.T. in Clinical Use*. New York: Grune and Stratton.
- <sup>2</sup> VERNON, P. E. (1936). The matching method applied to investigations of personality. *Psychol. Bull.*, 33, 149-177.
- <sup>3</sup> BURT, C. (1945). The assessment of personality. *Brit. J. Educ. Psychol.*, 15.
- <sup>4</sup> JACKSON, L. (1952). *A Test of Family Attitudes*. London: Methuen and Co.

### The Relation Between Knowledge of English Grammar and Ability in English Composition.

By NORA ROBINSON

(Abstract of thesis submitted for the degree of M.Ed. in the University of Manchester, 1959.)

#### AIM.

The aim of the enquiry was to consider the proposition that a knowledge of grammar is essential for the accurate and effective use of language and, in particular, to determine the degree of relationship between knowledge of grammar and ability in composition.



The specific aims were to establish the degree of association between the two variables, to investigate the effect of differences in age, sex and intelligence on the relationship and to determine whether a knowledge of grammar enables pupils to write English which is accurate, logical and of a good style.

#### METHODS.

Tests of grammatical knowledge, General Ability and composition were administered to all second year and fourth year pupils in four large, single-sex, urban grammar schools. Five pupils were chosen at random from each of the twenty-nine classes completing the tests and the resulting sample of 145 pupils was used for the investigation.

The intelligence test used in the enquiry was the Manchester General Ability Test, Senior I which has a reliability coefficient of .95. For the test of grammatical knowledge, eight ten-minute objective tests were constructed, one for each of seven parts of speech and one for sentence analysis; the coefficient of reliability for the eight tests was .96. The test in composition consisted of three thirty-minute essays, each of a different type.

Three marks were obtained from the essays: general accuracy in spelling, punctuation and grammar; grammatical accuracy alone; and general composition ability. The former scores were obtained by deducting points for errors; the latter score was derived from the marks of three examiners marking by the method of general impression. Coefficients of correlation between the pupils' three essay marks (general composition ability) were .569, .446 and .605; the intercorrelations of the marks of the three examiners were .891, .770 and .641. A random sample of ninety scripts was re-marked by the examiners and coefficients for mark-re-mark correlations were .93, .88 and .897. Lastly, the aggregate scores (general composition ability) for the first occasion of marking were correlated with the aggregate scores for the second occasion and a coefficient of .94 was recorded.

#### RESULTS.

##### COEFFICIENTS OF CORRELATION: GRAMMAR AND COMPOSITION.

				SCHOOLS			
		Years		Girls		Boys	
	Total	4	2	H	L	B	N
N=	145	75	70	35	40	40	30
1.	.572	.459	.369	.712	.441	.684	.357
a.	.490	.361	.363	.687	.259	.634	.223
2.	.493	.342	.508	.497	.502	.516	.409
a.	.447	.278	.505	.455	.426	.482	.351
3.	.307	.243	.293	.334	.332	.384	.211
a.	.251	.174	.281	.327	.328	.316	.172

1=general composition ability.

2=general accuracy.

3=grammatical accuracy.

a—first coefficient.

b=coefficient with intelligence held constant.

The calculation of coefficients of partial correlation indicates that interaction of intelligence with the two variables causes but a slight increase in the degree of association.

In general, there is a low degree of relationship between knowledge of grammar and the three aspects of composition examined in the enquiry. Of the three, coefficients are lowest for correlation between knowledge of grammar and grammatical accuracy in composition. Inspection of the scores awarded for grammatical accuracy reveals that these are high; pupils make few errors of this kind. There is a closer degree of relationship between knowledge of grammar and general composition ability in schools H and B than in schools L and N, and there is reason to believe that this is associated with more intensive teaching, especially of grammar. The effect of differences in age and sex on these conclusions appears to be slight.

Analysis of variance techniques were applied to scores for grammar, General Ability and general composition ability. There were significant differences, at the .1 per cent. level, between the second and fourth year groups for all three scores. There were no significant differences between the mean scores of boys and girls for any of the tests. For General Ability and composition there were no differences in the mean achievement of the several schools; for the grammar scores there were differences significant at the 5 per cent. level between schools N and L and schools N and B.

Perhaps the true test of the value of grammar lies not in the degree of its relationship with composition but in the nature of that relationship: is it effective? In the present enquiry there is no evidence that pupils from schools where there is a higher degree of association between knowledge of grammar and ability in composition, or pupils from schools with a significantly different level of attainment in grammar, obtain, as a group, marks in composition which are significantly different from those obtained by pupils from schools with neither of these characteristics.

## NEW TESTS

WISEMAN, S., and WRIGLEY, J. (1959). *The Manchester Tests: General Ability Test (Senior) 2; Reading Comprehension Test (Senior) 1; Mechanical Arithmetic Test (Senior) 1*. University of London Press. Manuals 1s. 6d. each. Test Blanks 36s. 2d. per hundred (intelligence and reading), 54s. 3d. (arithmetic); 6d. and 9d. each.

The authors point out that there is a dearth of standardised attainment and intelligence tests for 14+ pupils. These ones are excellently constructed, easy to give and score, and are standardised on a complete year group of over 13,000 pupils in Manchester schools (omitting independent, direct grant and E.S.N.). Deviation quotients run from 70 or below at 13:6 to 140 or above at 15:2. Presumably, the figures below 14:0 and above 15:0 are extrapolated. Split half reliabilities are between .95 and .97. Time limits are 60, 45 and 60 minutes respectively.

The intelligence test is in omnibus form, with 100 items, 48 of which are of the conventional verbal types: 13 are letter series, and as many as 39 involve numerical insight—perhaps rather an odd weighting. None are non-verbal.

The reading comprehension passages (eleven in number) are varied and well chosen for modern and grammar school pupils. Boys do a little better than girls, and there are separate sex norms for this and Arithmetic. The suitability of the content of Arithmetic may be more open to question, though admittedly, it is called 'mechanical' and may well be fairly typical of what is taught in modern schools. Nearly half the questions are of the type included in Part I of an 11+ test, and there are scarcely any verbally stated problems such as make up Part II; instead, the remainder are mostly fractions, decimals, percentages, and algebraic substitution.

The authors seem to have in mind the future publication of parallel tests. Perhaps they could help still further by providing non-verbal intelligence, English usage, Mathematics, and other tests.

EYSENCK, H. J. (1959). *The Maudsley Personality Inventory*. University of London Press. Manual 2s. 6d.; questionnaires 30s. 2d. per 100, 5d. each; scoring keys 2s.

This is a standard type of personality inventory, yielding scores for Neuroticism and Extraversion. A good deal more care seems to have been taken than in many American tests to secure a small number of items (forty-eight in all) which yield reasonable internal consistency and which avoid overlap between the two traits. Here the intercorrelation is usually around .15, probably because Extraversion is interpreted (like Guilford's rathymia) more as liveliness than as sociability. A short scale of twelve items is also provided, for which fair reliability is claimed.

Norms are based on a 'quota sample' of 1,800 normal adults, and results are quoted for various groups of hospital patients. It seems a pity that percentiles are not given, in addition to means and standard deviations. Professor Eysenck is commendably cautious in stating that this is to be regarded as a research instrument, designed for experimental and other purposes. Obviously it would be most undesirable if personnel selectors and others without adequate psychological training applied it wholesale, and interpreted its results as giving complete measures of whatever they thought neuroticism and extraversion to mean.

STOTT, D. H. (1958). *The Bristol Social Adjustment Guides*. University of London Press. Manual—*The Social Adjustment of Children*, pp. 46., 25s. Specimen set of blanks, scoring templates and summary charts, 9s. 6d.

Dr Stott has been working on his method of personality diagnosis for several years, and after numerous revisions, has now provided forms for the assessment of: (1) boys at school; (2) girls at school; (3) a child in residential care; and (4) the child in the family. The Manual, though somewhat difficult to follow, is worthy of close attention from child psychologists and others interested in personality. 11r.

Stott criticises both the rigidity of the 'objectivist,' with his fixed traits or types of personality, and the subjectivity of the clinician. All that he claims to supply is a series of records of behaviour in particular contexts, grouped under a number of syndromes or 'attitudes' (Anxiety, Hostility to Adults, Restlessness, etc.) and graded for their seriousness in respect of unsettledness or maladjustment. Clearly, these can be of considerable value to the psychologist in the educational clinic, the teacher, or others concerned with maladjusted or delinquent children, in giving a codified, all-round picture of a child's personality. They represent a genuine attempt to get away from the artificiality of the objective, projective or questionnaire test, and the bias of the ordinary personality rating.

Nevertheless, a number of doubts arise. First, how far can any observer, who has sufficient acquaintance with the child to answer the questions, avoid halo and subjectivity even in underlining such apparently concrete 'bits of behaviour'? Dr. Stott provides some evidence of reliability (inter-observer agreement), but other studies of the problem reported in the literature are unpromising.

Secondly, it is not easy to follow just why these particular categories were chosen, nor how the items were assigned to one rather than another. Quite legitimately, the author eschews any overall score or quotient for maladjustment, but one would like to know much more about the internal consistency of the categories and their overlapping. Thirdly, while frequency tables for the 'School' symptoms are provided, it would be valuable to have something in the nature of norms for each category, and for different age groups. It seems unlikely that the same patterns of normality and abnormality occur in children all the way from 5 to 15 years. However, this type of information can well be collected by the author and others who use the Guides, and we will look forward to future reports.

It is unfortunate that the cost, particularly of the Manual, is so high. Packets of each of the four rating forms (the 'Guides'), and the summary sheets ('Diagnostic Forms') can be obtained separately.

SOUTHGATE, V. (1959). *Southgate Group Reading Tests*, Forms A, B and C. University of London Press, price 5d. per test blank, Manual 2s. 6d.

Miss Southgate has done a useful job in making available a group reading test over the 5:9 to 7:9 age range, in three parallel forms. It follows the conventional word-recognition form: that is, each child has to ring one word, read out by the tester, in a group of five. The test is untimed, and takes about 15-20 minutes. The blanks are well-printed and the Manual clear.

It is a pity, perhaps, that no information is given on the effects of children's familiarity with the tester's pronunciation and of the clarity of his enunciation. The high reliability quoted as between different testers is meaningless, since it is based on a wide age range. However, about half the items are accompanied by pictures; and the correlations of around 0.90 (in a 6-year age group) with individual oral tests, are reassuring. With reading comprehension, as might be expected, the correlation appears to be much lower though the quoted figure—is again—meaningless.

The thirty items in each form provide fair discrimination, namely about one to two months per item. Standardisation was based on over two thousand children in the City of Worcester. One wonders whether the assumption that this sample is representative of the country is legitimate, but it had the advantage of being a complete population. The author might also have provided percentiles at successive age levels. She claims, reasonably enough, that teachers do not understand standard scores; but one fears that the absence of indications of spread mean that there is little discrimination among below-average younger, or above-average older, readers.

P. E. VERNON.



## BOOK REVIEWS

BARANYAI, E., and LENART, E. (1959). *Studies in the Psychology of Thinking in Written Communication*. (In Hungarian.) Publication of the Hungarian Academy of Sciences, Budapest. p. 316.

This is a study of simple communication in children's composition. Its object is to throw light on the psychology and development of thinking, with particular reference to logical thinking and forms of communication. The investigation was carried out on school children between the ages of 10-14 from General Schools in Budapest. Two different methods were used :

(1) Various types of composition were devised for the purpose of the study, such as preparing a sketch from a given essay ; repeating the pattern of a given essay in an essay on another subject ; writing an introduction and conclusion to a given piece of writing ; arranging the cut-up paragraphs of an essay in a logical order ; fitting a sentence lifted from its context into its correct place, and completing sentences.

(2) Analysing children's compositions written in the course of Hungarian lessons at school. Two groups of essays were analysed : (a) free compositions ; (b) essays written on subjects requested by the investigators. The main points investigated were : logical sequence, correct reasoning, and structure, i.e., relationship between parts and the whole, and the use of direct and indirect speech.

The exhaustive and competent analysis of the material with ample quotations and illustrations rather bears out our knowledge on the subject than throws new light on it. The main factor in the development of logical thinking and abstract reasoning is age, and the trend of development leads from the concrete to the abstract as demonstrated by Piaget.

The outstanding merit of the study is the contribution it makes to the teaching of composition and to the use of composition as a means of developing logical thinking. It is shown in great detail where and why children fail in satisfying the demands of logical thinking, but it also demonstrates the weaknesses of teaching methods, for example, the over-emphasis on personal experience in collecting material and the corresponding lack of experience in objective, detached description.

M. VIDOR.

This is an interesting and stimulating study.

BECK, S. J., and MOLISH, H. B. (edit.) (1959). *Reflexes to Intelligence: A Reader in Clinical Psychology*. Illinois : Free Press of Glencoe, pp. xiv + 669, \$8.50.

This is a collection of seventy-three articles tracing the development of clinical psychology. The papers, which are taken from various sources, are unabridged and over two-thousand references are given. The quality of the selection is excellent, though there are some notable omissions, and the very names of the authors indicate the worth of the individual contributions. The major viewpoints are represented and the articles are separated into six sections each preceded by an original contribution from the editors.

As an introductory reader in clinical psychology it seems too lengthy and difficult. The numerous details of so many research articles tend to obscure the main trends and the editors' introductions do little to remedy this. Treatment is omitted and there is no index.

To whom will this book be useful ? For it should be useful, containing as it does such a large number of important articles. Probably to those who are already familiar with the field, to thesis writers and to lecturers who want original sources and detailed reference lists. To the practising clinical or educational psychologist seeking a clear exposition of current trends it is of limited value.

MAURICE J. TYERMAN.

Buros, O. K. (edit.) (1959). *The Fifth Mental Measurements Yearbook*. New Jersey: Gryphon Press, pp.xxvii+1292, \$22.50.

This is the eighth publication of a series which began twenty-four years ago in order "to assist test users of all kinds—educators, teachers, personnel workers, psychiatrists, psychologists, sociologists and others—in locating and evaluating tests and books on testing." That is still the fundamental purpose. By emphasis on sound, unbiased reporting of tests and reviews, the editor claims that the year-books have helped test users, constructors, and reviewers to be more critical and discriminating.

The Fifth Yearbook covers a seven-year period from 1952 to 1958. The main two sections deal with tests and books. In all, 957 tests are listed together with 746 reviews of tests; there are 6,468 references on the construction, use and limitations of specific tests. Altogether 485 books on mental measurement and allied subjects are listed together with 535 excerpts from reviews of these books. There are classified indexes of publishers, titles, and authors and a most useful index of tests listed according to subjects. More attention is given to the "sections covering achievement batteries, non-projective personality tests, English, foreign languages, group intelligence tests, mathematics, miscellaneous areas, reading, science and social studies."

The general plan follows the pattern of previous yearbooks and cross references are easily traced in this and earlier editions. Brief information is supplied, where available, on essential test facts: age range, method of scoring, time of test and subtests, cost and when possible reliability and validity figures; too often these last two important facts are not supplied by the test constructor, but critical and pertinent reviews often reveal why this information was withheld.

This edition gives comprehensive and up-to-date information on test material published in English-speaking countries, but it is sad to note that because of increased costs, and the vast growth of material on mental measurement, this may well be the last of these valuable reference volumes. It is proposed that the Sixth Yearbook should cover work in the United States. A smaller book, *Tests in Print*, published at more frequent intervals, would contain a comprehensive bibliography of titles published in English-speaking countries. If this is the last 'Buros' with its comprehensive and near-international coverage on mental measurement, it is most fitting that it should have been dedicated to Sir Cyril Burt.

W. H. KING.

CLARK, M. M. (1959). *Teaching Left-Handed Children*. Scottish Council for Research in Education, Publication No. XLIV. University of London Press, pp. 44, 2s. 6d.

This pamphlet aims to make available to teachers and others a simple statement of the findings of Dr. Clark's recent book on *Lefthandedness*. But, in fact, it also covers further useful ground in giving rather detailed guidance on teaching the left-handed child to write. As Dr. Clark points out, most of the 5 to 8 per cent. of left-handed writers existing in our schools will have suffered frustrating experiences through ill-judged attempts to get them to write with the right-hand, or through mere neglect if the left hand is tolerated; and there must also be many more who have managed to change, but who might have written with greater facility and less emotional disturbance, had skilled instruction and understanding been available.

CRONBACH, L. J. (1960). *Essentials of Psychological Testing* (2nd ed.). New York: Harper, pp.xxi+650, \$5.00.

Many connoisseurs of mental testing considered the first (1949) edition of Cronbach's book to be the most stimulating and informative of any in this field. The new edition is bigger and better, though perhaps a little less systematic in its treatment. It is very up-to-date, over two-thirds of the references having been published since the previous edition; and British work receives a good deal more attention than in most American texts. However, its very wealth of penetrating

ideas makes it rather less suitable as a textbook than some of its more pedantic competitors. Rather it should be regarded as advanced reading for the sophisticated worker in guidance and selection. At the same time it is extremely readable; for example, the simple interpretations of factor analysis and decision theory are real *tours de force*.

The main sections are: Purposes and types of tests; Administrative procedures; Types of scoring; Validity and reliability; Individual tests of general ability; Group tests; Classification of abilities, differential testing, tests of special abilities; Personnel selection; Educational attainments; Tests of interests; Personality inventories, ratings, behavioral tests and projective techniques; Clinical approaches. P.E.V.

DEESE, JAMES (1958). *The Psychology of Learning*. London: McGraw-Hill, pp.vii+367, 50s. 6d.

This is primarily intended as a text book for psychology students and seems to the reviewer one of the better things of its kind. It is closely packed with useful material; yet its lucid style and clear arrangement make it easy to read. It is divided conventionally into chapters on different aspects of learning, but more care than usual has been taken to connect the subject matter of different chapters, and the student should find it easier to unify what he learns.

The book as a whole aims to give an outline of the present state of learning psychology. The author points out that the period of comprehensive explanatory theories and controversial warfare between rival 'schools' is now largely over, at least in the experimental branches of psychology. He is not, himself, attached to any one broad theory. The student is, rather, invited to consider more limited theoretical questions; for instance, the nature of reinforcement; the function of verbal mediation in problem solving, and so on. The present state of the evidence is summarised and the author discusses what can be taken as established, what may be provisionally accepted on the balance of the evidence, and what is still uncertain.

The main emphasis of the book is on conditioning, but insight, hypotheses and problem solving are also discussed. There are also useful chapters on motivation and emotion, in connection with learning. A final chapter on "Theory and Applications of Learning," contains a rather curious assortment of topics. Association *versus* Gestalt theory, neurophysiology, imprinting, psychotherapy, socialisation and education are treated in rapid succession, and rather too briefly to be very useful. But the book, as a whole, should be a considerable help to students and of interest to many others who are not specialists in learning psychology, but wish to keep themselves informed about it.

S. C. RICHARDSON.

EDWARDS, A. (1958). *Statistical Analysis* (Rev. edit.). New York: Rinehart, pp. xiii+234, \$4.00. *Workbook*, pp. 76, \$1.25; Text and Workbook, \$5.00.

This is a sound, conservative, statistical text which should provide the student with a thorough understanding of basic principles. It does not attempt to go beyond the standard topics (for example, elementary analysis of variance only is included, and factor analysis is excluded), but it makes the why and wherefore of statistical techniques unusually clear. It also makes little concession to practical applications, its examples being chosen, rather, with a view to ease of calculations. It is accompanied by a Workbook containing some very testing questions on each chapter which are not, be it noted, in multiple-choice form; the student really has to work to answer them. Presumably, the psychology courses in our own universities and training colleges are too variegated for British authors to find it worthwhile issuing similar Workbooks. But they could be a valuable teaching adjunct.



EDWARDS, A. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: Holt, pp. xv + 108, \$3.25.

The halo effect in ratings and self-ratings, and the susceptibility of personality inventories to faking, have long been recognised. Professor Edwards shows that, by getting judges to assess inventory items according to social desirability, very high correlations (about 0.85) are obtained with the frequency of endorsement of those items. Thus, the high correlations so often found between neuroticism and introversion tests, or between the various MMPI scales, can be largely accounted for in terms of the tendency of testees to endorse socially desirable times. What the author does not try to answer is whether this social desirability tendency is merely a kind of 'response set' or a meaningful personality characteristic. Probably it is both, in much the same way as halo in ratings partly represents a genuine tendency for 'good' character traits to inter-correlate, partly a tendency to rate people we like high on all desirable traits. Nevertheless, Edward's systematisation of this topic will be of value to all who work with inventories, ratings, Q-sorts and the like. Moreover, as his own *Personal Preference Schedule* shows, it opens up possibilities of forced-choice tests or ratings which are less susceptible to these distortions.

FRASER, E. (1959). *Home Environment and the School*. Scottish Council for Research in Education, Vol. XLIII. University of London Press, pp. vii + 83, 10s. 6d.

The Scottish Research Council has performed a useful service to educational psychologists in publishing Miss Fraser's investigation, which was carried out some 10 years ago. She performed the astonishing feat of interviewing, in their own homes, a parent or parents of each of 408 Aberdeen pupils, in order to assess aspects of home environment relevant to school achievement in the first year of the secondary school. Parents' own education, book and newspaper quality, income per person, father's occupation, family size, rooms per person, attitudes to education, broken or abnormal home and general home atmosphere (rating) all gave substantial correlations with Moray House I.Q., and slightly larger ones with present attainment. Mother at work did not. In combination these factors yielded multiple correlations of .687 and .750, respectively, the most efficient predictors being income, attitude and abnormal home.

Miss Fraser provides a sensible discussion of the relations of intelligence and education, and how environment affects them, and she brings out the important educational implications of her findings.

FRIEDENHAIN, P. (1959). *Write and Reveal: Interpretation of Handwriting*. London: Owen, pp. 183, 30s.

Like many recent books on graphology, this one protests that specific features of the script are no longer supposed to have fixed significance for personality, and that modern graphology—based on Saudek's principles—is an exact and reliable method; and then proceeds to draw large numbers of conclusions from signs, or from intuitive guesses, and to validate these merely by presenting specimen analyses which somebody certifies to be a faithful picture of the writer.

This example is well written and attractively put together, with numerous illustrations. It includes a number of interesting pathological cases, and of instances where the graphological diagnosis is said to have helped in vocational guidance. One's doubts are, if anything, enhanced by the claim that the scripts of children as young as 9 years provide valuable material for educational guidance. Fortunately, the author does not encourage amateurs to undertake graphological analyses on their own, and she stresses the need for further research.

P.E.V.



GRONLUND, N. E. (1959). *Sociometry in the Classroom*. New York : Harper and Brothers, pp. xviii+340, \$4.50.

Since Moreno published *Who Shall Survive* in 1934, there has been a growing volume of sociometric research. Most of this has been reported only in specialised journals and the general reader has had to be content with passing references in books of a more general nature. *Sociometry in the Classroom* is to be welcomed as remedying this situation. The book contains an account of sociometric techniques with directions for giving and recording the results of sociometric tests, a discussion of the meaning of the results, and suggestions for applying them in the solution of educational problems.

Too often sociometric writers have been more noted for their enthusiasm than for their powers of criticism. The present author is to be commended for showing both the importance of the technique and its limitations ; for it must be admitted that the very ease with which a sociometric test can be applied has induced what can only be described as 'test happiness' in many devotees. This sane and judicious account of the ways in which the results of a sociometric test can add to our knowledge of a group of children and give guidance for their future treatment should do much to ensure that those who adopt the methods described do so with a full appreciation of what the test does and does not tell us.

If the book has a weakness, it is the fewness of the sociometric diagrams shown. A large part of the appeal of sociometry has always been the clarity with which a sociogram portrays the relationships within a group of people. A book which sets out to describe the basic techniques of sociometry should surely base more of its exposition on a study of sociograms than is done here. This, however, is a minor point, and the book can be strongly recommended as satisfying a real need. It should find a place in all training college libraries.

K. M. EVANS.

HILGARD, ERNEST R. (1958). *Theories of Learning*. (Rev. ed.) London : Methuen, pp. vii+563, 45s.

The revised edition of Hilgard's *Theories of Learning* contains most of the earlier book. Two of the original chapters, those on Wheeler and Maier, are omitted ; other changes are almost entirely in the form of additions. Chapters on some earlier writers, such as Guthrie and Skinner, have been brought up to date by reference to recent work and the one on Hull has been considerably extended. The author's evaluation of the theories discussed is not greatly changed. The new parts of the book are a chapter on Freud ; another on mathematical models ; and two chapters on 'current developments'. These last chapters are concerned mainly with theories of reinforcement and discrimination learning, and with neurophysiological models.

In surveying the present state of learning theory, the author maintains his own eclectic position. He suggests that differences between learning theories are still partly due to the fact that experimenters tend to concentrate on one type of learning experiment. He maintains his earlier view that we should be prepared to find different types of learning process at different evolutionary levels, and that some types of human learning may be qualitatively different from learning in animals.

Non-American readers may feel that there is rather little mention of work done outside the United States. Even Pavlov receives only two passing references. But the book remains the standard account of the search for a satisfactory theory of learning, and the revised edition will be welcomed by most people interested in the history of learning theory and its present and future development.

S. C. RICHARDSON.

HUGHES, A. G., and E. H. (1959). *Learning and Teaching*. (3rd ed.) London : Longmans, Green, pp. x+484, 14s.

The third edition of this 'classic' text-book is sub-titled, as were the previous editions, 'An Introduction to Psychology and Education'. It differs from them in having been, in part, re-written in an endeavour to take account of some of the recent changes in education and some of the recent developments in educational

psychology. The authors have also undertaken a partial revision of the bibliographies, which follow most chapters and they have added numerous, and sometimes lengthy footnotes. Their convictions remain substantially the same as they were in earlier editions as does their sincere concern for education.

It is a formidable task for any pair of authors to undertake to integrate into a text-book two decades of educational research and change. That the authors have not entirely succeeded is no reflection on their efforts. It is rather a reflection on the magnitude of the task.

P. H. TAYLOR.

HUSEN, T., and HENRYSSON, S. (1959). *Differentiation and Guidance in the Comprehensive School*. Stockholm: Almqvist and Wiksell, pp. 196, no price quoted.

Since 1950, Sweden has been introducing education for all pupils between 7 and 16 years in common schools. This is a report of a Conference arranged in 1958 by the Swedish Ministry of Education for the discussion of school organisation (streaming, specialisation, etc.) and the educational development and guidance of the adolescent. It was attended by experts from sixteen countries, and the accounts of their own plans (all translated into English) are of interest to the British reader, particularly since the participants concluded that gradual differentiation and guidance from 11 to 16 is much to be preferred to selection at 11+. Another specially noteworthy point is that the Swedish Ministry is co-operating with the Stockholm Institute of Education in scientific research on the development of interests and abilities and the extent to which they differentiate over this period. It is difficult to imagine our own administrators basing their policies in the same way on the results of research.

MORTENSEN, D. G., and SCHMULLER, A. M. (1959). *Guidance in Today's Schools*. New York: Wiley; London: Chapman and Hall, pp. viii + 436, 46s.

The detailed description of counselling techniques, the twenty-one specimen forms for use by counsellors and the illustrative case-histories are useful practical aspects of this manual. The setting is American: of 380 references only five are British, two mis-spelt. The authors maintain that a book on guidance must include "background materials relating to the philosophical and scientific bases of guidance." Consequently, the first half of the book discusses the meaning of guidance and its theoretical basis. Almost every relevant aspect of psychology receives passing mention, four pages to Freud, two to behaviourism, three to sociometry, and so on. This broad shallow treatment of the psychological background to guidance may indeed reduce the appropriateness of the book to Britain, where guidance is not a separate specialism and the teacher or psychologist concerned with guidance will (or should) have learned this background from wider and more systematic study. The sceptic, suspicious of the separation of the roles of counsellor and teacher, will not be reassured by the implied superficiality of counsellors' educational and psychological knowledge. However, used selectively, the text provides valuable practical suggestions for those who are prepared to recognize the importance of this aspect of education.

JOHN NISBET.

NATIONAL ASSOCIATION FOR MENTAL HEALTH (1959). *Truancy—Or School Phobia?* Proceedings of the 15th Inter-Clinic Conference. London: National Association for Mental Health, pp. 40, 3s 6d.

In Britain, until a few years ago, absence from school was considered a minor offence: it is now seen as an important and complex symptom. The report of the 1958 N.A.M.H. Conference emphasises its significance, and while containing little that is new, effectively summarises most of the basic facts and gives suggestions as to treatment. It stresses the point that persistent truancy and school phobia shade

into each other and that children showing these symptoms tend to be lonely, insecure and unhappy. It is also made clear that in nearly every case the main trouble is not in the school but in the home, and that treatment must be based on meeting the child's individual needs.

In general the reports are untechnical, lucid and enlivened by first-hand descriptions. This is a pamphlet worth reading and keeping.

MAURICE J. TYERMAN.

OLSON, W. C. (1959). *Child Development*. Boston: Heath, pp. xi+497. No price given.

This is the second edition of a book first published in 1949. In it, the Dean of the School of Education at the University of Michigan gives to teachers and intending teachers a broad, well-documented and reasonably balanced review of earlier and up-to-date experimental findings in the field of child development. His book covers the topics of Physical, Mental and Emotional Development; the Growth of Educational Achievement; the Child as a Whole; Human Relations in the classroom; the Child in the Home and Community; the Affective Life of the Child; and the Teacher in Individual and Group Relationships.

One may fairly say that the author's fundamental belief is that the child grows as a whole. Indeed, arising out of the work of S. A. Courtis, B. O. Hughes and Olson himself, there has developed the concept of *Organismic Age*, which is derived from both physical and psychological measurements, and the author makes great use of this throughout his book. It is his clear view that learning is a function of total growth. Only future research can tell us exactly how valuable this concept of Organismic Age will be. It can be said, however, that he has drawn our attention to an aspect of learning in children that we sometimes overlook. For example, it is now manifest that the effects of remedial education with many children are not as satisfactory as was hoped ten years ago. Perhaps we have placed too much stress on the backward child's emotional life, sociological conditions, and teaching methods *per se*, and overlooked genetic and constitutional influences that cause children to mature at very different rates. In a way, Olson's concept is not altogether new; Burt used somewhat similar concepts in Chapter 6 of *The Backward Child*.

British readers may be disappointed to find so few references to relevant work done in this country. They will also be surprised to find that the work of Piaget is dismissed in eighteen lines. There are a few inaccuracies; the information (p. 45) that there are forty-eight chromosomes in most body cells and not forty-six, might be explained by new evidence coming to light between the time of the book going to press and publication; but British Courts have not been able to award corporal punishment to delinquents since 1948. However, the book is very readable and to be recommended.

K. LOVELL.

PIAGET, J. (1960). *Language and Thought of the Child*. (3rd English ed.) London: Routledge and Kegan Paul, pp. xxiv+286, 25s.

PIAGET, J., and INHELDER, B. (1959). *La genèse des structures logiques élémentaires*. Neuchâtel: Delachaux et Niestlé, pp. 295, fr.s.13.

Piaget's book on language, first published in 1923, reappears in its third English edition, with a new chapter devoted to a comparison of egocentrism in the language of young children when talking to an adult with that obtained in conversation with their peers. Piaget finds that about the age of three, egocentrism is greater in the company of an adult than in that of other children; by the age of four the proportions tend to be equal. Socialised remarks addressed to an adult show an overwhelming predominance of questions, a feature which is not apparent in socialised remarks occurring between children. Piaget sees support in this finding (based on the conversations of four children) for his contention that the decline of egocentrism owes much to the conflict of view-points and opinions occurring in the peer society—where the child does not adopt a frankly subordinate role. Replying to criticisms, Piaget is at pains to point out that the fundamental fact is intellectual egocentrism, while



language offers no more than a somewhat unreliable index which varies with the temperament of the child and with the character of his upbringing. The ideas contained in the work as a whole are more provocative, and perhaps more cogent than the observational and experimental evidence adduced in their support.

*La g nese des structures logiques  l mentaires* is a study of development in classification and seriation. The former especially, the discussion of which takes up eight out of the ten chapters in the book, would seem to be of particular significance to the development of thinking as a whole. An elementary classification may be recognised in the adaptive behaviour of quite young children in relation to objects, and later in even the simplest use of words. But true classification for Piaget involves at least an implicit recognition of the logic of classes. The test of such operational understanding is to be found in children's handling of hierarchical class structures and of cross-classificatory matrices. Both attain an operational level about the age of 7-8. The child of six who is faced with a collection of beads containing square red beads, round red beads and round blue beads will agree that not all the round beads are blue because he can see the red round beads, but by the same token he will deny that all the blue beads are round (because some round beads are red!). He can only compare the two classes if they are co-extensive. The difficulty, according to Piaget, is that of co-ordinating the extensive and intensive properties of classes. It may be that, psychologically, the essential difficulty is to consider the same objects in terms of two variables and co-ordinating the two judgements. In any case Piaget's insistence on a more rigorous analysis of classificatory behaviour than that demanded by mere sorting into disjunctive groups is fully justified. Logical reasoning depends on the ability to recognise similarities (e.g., within genera) and differences (e.g., between species), for only thus can the thinker free himself from a pre-logical argument by analogy. A noteworthy feature of this book is that the authors provide full quantitative tables to illustrate the findings of their several researches. The total number of children examined is 2,159.

The simultaneous appearance of these two volumes suggests some comparison between the earlier and later phases of Piaget's work. On the one hand, his methods are more rigorous and experiments are devised in a less ambiguous relation to the theoretical points at issue. On the other, these conclusions themselves often appear less general in their import. Again, where earlier Piaget stressed the lack of social experience to account for the absence of logical thinking, Piaget now emphasises the one-sidedness of perception. The latter approach often leads to an over emphasis on the indispensability of a concrete support to operational thinking below the formal level. That such an emphasis is incorrect is clear from the coincidence of the Terman-Merrill item on similarities and differences at the critical age of 7-8. A weakness in all of this work is the relative depreciation of the role of language itself in the development of thought. It remains that the earlier work of Piaget affords the most cogent picture which we possess of the kinds of ideas which animate the thinking of the young child, while his more recent studies offer a powerful tool whereby we may be able to analyse the logical and psychological difficulties in the elements of the school curriculum and gear our methods to the understanding of the child.

E. A. LUNZER.

REDL, F., and WATTENBERG, W. W. (1959). *Mental Hygiene in Teaching*. (2nd ed.) New York: Harcourt, Brace, pp. xiv+562, no price given.

As its title implies, this book is essentially a textbook for teachers who wish to be helped in their work by modern psychology. The book achieves its object, although its material is presented in a somewhat elementary manner. Overtly, an anecdotal and somewhat chatty approach is used. Covertly, the structure of the book reveals that the authors know not only their psychology, but also the many problems which confront novice teachers. The authors present with subtlety the pegs of psychology on which they hang popular clothes borrowed from divergent theoretical fields of general and clinical psychology. For instance, the earlier chapters are based on psycho-analysis (without calling it such) in an attempt to give the reader a fairly unified theory of Personality.



Part A, amplifies the introduction, Part B, appositely headed "Some Fundamentals," deals in five chapters with such topics as Motivation, Behaviour Mechanisms and Growth, etc. In Part C, the meat, though thin, begins to appear at last, and the scene of general psychology is switched to that of social psychology, without calling the reader's attention to it. It is headed "Classroom Applications," and subdivided into seven popular eye-catching titles such as: "Unusual Learners," "Group Life in the Classroom," etc. Here one encounters, in disguise, such well weathered friends as Moreno and his sociogram, and Lewin, Lippitt and White, and their classical social climate experiment. Newer and more personal ground is broken in later chapters which deal with the various roles that the teacher has to play in relation to his pupils, where he has to act alternately as 'trouble-shooter,' as parent substitute or as target for hostile feelings. He is also given advice as to how to deal 'psychologically' with complex classroom situations, such as arise when a boy has stolen another's money, etc. The chapter called "Influence Techniques" deals with problems affecting discipline and leadership at school, where the teacher is advised not to enforce discipline crudely, but to devise methods which support children's own self-control via a display of 'humor,' by closing in rapidly on a potential mischief maker, or by removing "seductive objects" (inkwells are not mentioned). Incidentally, the time-honoured traditional English Public School stick finds no favour in this American textbook. Teachers are advised to give naughty children a 'chance of cooling-off' in corners, and are warned of the dangers of instituting 'mass-punishment' in order to detect a single culprit. The last chapter advises teachers who are in conflict to stick to the rule of the golden mean in most circumstances affecting discipline or their own anger when aroused. The authors thus show that a solution in conflict situations is possible (despite their gaffe in the chapter heading, which includes in this connection, the word 'dilemma').

The last section of this book, called 'Special Problems,' is in some ways the most interesting one, for here the thread of 'general psychology' changes to that of 'psychopathology.' Interest derives less from the factual discussion and advice on the detection and management of neurotics, delinquents, geniuses and dullards, than from a spot-light on the psychological make-up of the teacher himself. Here are real glimpses of depth-psychology. Readers are made aware of the necessity for teachers to face up to some form of an inner self-analysis, especially in regard to their vocation. May one have taken up teaching to satisfy such needs as: status gain; love for a subject; identification with one's own former teacher; love for children, or because of a reaction against one's own unsatisfactory childhood in an endeavour to create an easier world for the new generation; the desire to re-live childhood patterns, or because of a pre-conscious craving for affection or power?

In recognising some of these variables, it is indirectly suggested that a cathartic effect takes place within the teacher, which in turn, must have a stabilising influence on a class. Teachers are also advised to learn that experience will ultimately help them to cope with complex problems, and that in personally stressful situations they may be aided by taking up a hobby, or by consulting a colleague—or even a psychiatrist!

In a final chapter, the authors refer wisely and modestly to the "Limitations of Mental Hygiene in Education," and thereby, as experts, let the cat out of the bag in front of their essentially non-expert readers in the field of psychology. They perform this task of deflation gently and with agility, covering themselves impressively by supplying a list of current references (including audio-visual ones) at the end of each chapter, and by attaching an appendix on sources of the literature and agencies in this field, and the meaning of special terms, all for the unhibited use of budding teachers and psychological greenhorns. The psychological soup of this book has been well watered down in its approach to provide a primer in a series which could well be called "Mental Hygiene without Tears." It is, nevertheless a useful textbook, especially for newcomers and potential converts.

R. G. ANDRY.

TERMAN, L. M., and ODEN, M. H. (1959). *Genetic Studies of Genius, Vol. I—The Gifted Group at Mid-Life*. Stanford University Press, pp. xv+187, 36s.

This volume carries the follow-up of Terman's gifted group to nearly thirty-five years after the initial testing, the mean age being now 44. The information was collected and analysed before Terman's death, but most of the writing-up is the work of his co-author. Nearly 1,300 subjects, representing 93–95 per cent. of the original group are covered.

The present volume is perhaps the least interesting of the five, since it adds little to the findings of Vol. IV. One gets the impression, also, that the interpretation of the data is less skilled and lively than it might have been in Terman's own hands. For example, the results on present intelligence consist of the scores of the gifted group, their spouses and other high-grade groups on a new form of the Concept-Mastery test, together with Terman-Merrill I.Qs. of 1,525 of the gifted population norms (except through the fact that the gifted group have significantly improved on their scores of 10 years ago); and the latter is almost equally meaningless—for while the average child I.Q. is 132.7 (one third scoring 140+ and only 2½ per cent. below 100), no attempt is made to relate these figures to the Standard Deviations of I.Qs. at the ages of testing.

The record of physical and mental health is as good as ever. It is shown that nearly 70 per cent. have graduated from college, as against 12 per cent. in the general population, and that very large numbers have proceeded to higher degrees or academic distinctions. The main alteration in the past ten years is the rise in occupational status, 86.3 per cent. instead of 71.1 per cent. now falling in the top two occupational classes, and only 1.2 per cent. as against 6.9 per cent. in the semi-skilled class or below. As a whole, also, the group has become noticeably more conservative in its attitudes.

*Genetic Studies of Genius* remains the most impressive demonstration of the favourable prospects of a child with high I.Q., and a monument to Terman's life work. It is good to know that this work will be carried on and that further volumes on the later careers of the subjects are intended.

P.E.V.

TRIST, E. L., and SOFER, C. (1959). *Explorations in Group Relations*. Leicester University Press, p. 68, 7s. 6d.

The first full-scale experiment in Britain with the laboratory method of training in group relations was held in September, 1957, by the University of Leicester and the Tavistock Institute of Human Relations, and was attended by forty-five members, roughly equally divided between industrial and non-industrial occupations. Through participation in study and application groups, the members were led to an understanding of some of the psychological forces which operate in such groups. This report outlines the methods and results of the conference, but gives little detailed information. There is, however, a useful bibliography which will guide those who wish to pursue the subject further. The method might be of value in the training of teachers, but would need to be used with caution and only by those who fully understood its implications and were capable of dealing with the conflicts which might be aroused.

K. M. EVANS.

### PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

J. W. THIBAUT and H. H. KELLEY (1959). *The Social Psychology of Groups*. New York. John Wiley, London: Chapman and Hall, pp. xiii+313, 56s.  
P. HALMOS (edit.). *Papers on the Teaching of Personality Development*. Sociological Review Monograph No. 2. Keele: University College of North Staffordshire, pp. 149, 12s.

- A. L. EDWARDS (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden Press, pp. xiv+108, \$3.25.
- A. L. EDWARDS (1959). *Statistical Analysis* (rev. edit.). New York: Rinehart, pp. xii+234, \$4.00. Accompanying Workbook, pp. 76, \$1.25; Textbook and Workbook, \$5.00.
- L. POSTMAN and D. A. RILEY (1959). *Degree of Learning and Interserial Interference in Retention*. Univ. California Publ. Psychol., Vol. 8, No. 4, pp. 271-396, \$2.50.
- M. READ (1959). *Children of their Fathers: Growing up among the Ngomi of Nyasaland*. London: Methuen, pp. 176, 18s.
- H. H. SOMERS (1959). *Analyse Mathématique du Langage: Lois Générales et Mesures Statistiques*. Louvain: Nauwelaerts, pp. 40, FB.80.
- P. FRIEDENHAIN (1959). *Write and Reveal: Interpretation of Handwriting*. London: Owen, pp. 184, 30s.
- M. M. CLARK (1959). *Teaching Left-handed Children*. Scottish Council for Research in Education, Publ. No. XLIV. University of London Press, pp. 44, 2s. 6d.
- I. R. WECHSLER and J. REISEL (1959). *Inside a Sensitivity Training Group*. University of California, Los Angeles: Institute of Industrial Relations, Monogr. Ser. No. 4, pp. 133, \$2.00.
- R. S. PETERS (1959). *Authority, Responsibility and Education*. London: Allen and Unwin, pp. 137, 12s. 6d.
- Z. P. DIENES (1959). *Concept Formation and Personality*. Leicester University Press, pp. xi+82, 18s.
- E. GALANTER (edit.) (1959). *Automatic Teaching*. New York: John Wiley; London: Chapman and Hall, pp. viii+198, 26s.
- A. V. JUDGES (edit.) (1959). *The Function of Teaching*. London: Faber and Faber, pp. 172, 15s.
- J. PIAGET (1960). *Language and Thought of the Child* (3rd edit.). London: Routledge and Kegan Paul, pp. xxiv+286, 25s.
- F. V. SMITH (1960). *Explanation of Human Behaviour* (2nd edit.). Constable, pp. xi+460, 30s.
- V. SOUTHGATE (1959). *Southgate Group Reading Tests, Test 1* (Forms A, B, C) Word Selection. University of London Press, 5d. each test blank, Manual 2s. 6d.
- L. TUSSING (1959). *Psychology for Better Living*. New York: John Wiley; London: Chapman and Hall, pp. xii+496, 48s.
- T. HUSEN and S. HENRYSSON (1958). *Differentiation and Guidance in the Comprehensive School* (Report on the Sigtuna Course, 1958). Stockholm: Almqvist and Wiksell, pp. 196, no price given.
- WORLD FEDERATION FOR MENTAL HEALTH (1959). *Africa: Social Change and Mental Health*. London: W. F. M. H., pp. 40, 3s. 6d.
- D. H. FUNKENSTEIN (ed.) (1959). *The Student and Mental Health: An International View* (Proceedings of Princeton Conference, 1956). Cambridge, Mass.: Riverside Press, pp. xxv+495, 35s.
- F. J. McDONALD (1959). *Educational Psychology*. San Francisco, Calif.: Wadsworth Publ. Co., pp. xix+748, price not given.
- R. KNIGHT and M. KNIGHT (1960). *A Modern Introduction to Psychology* (6th edit. rev.). University Tutorial Press, pp. 272, 9s.
- B. N. PHILLIPS, R. L. DUKE and M. V. DE VAULT (1959). *Psychology at Work in the Elementary School Classroom*. New York: Harper, pp. xii+395, \$5.00.
- S. L. PRESSEY, F. P. ROBINSON and J. E. HORROCKS (1959). *Psychology in Education*. New York: Harper, pp. xi+658, \$6.50.
- L. J. CRONBACH (1959). *Essentials of Psychological Testing* (2nd edit.). New York: Harper, pp. xxi+650, \$5.00.
- G. MURPHY (1958). *Human Potentialities*. New York: Basic Books. London: Allen and Unwin (1960), pp. x+349, 25s.
- INCORPORATED ASSOCIATION OF ASSISTANT MASTERS IN SECONDARY SCHOOLS (1960). *Teaching in Comprehensive Schools*. Cambridge University Press, pp. 48, 2s. 6d.
- M. L. J. ABERCROMBIE (1960). *The Anatomy of Judgment*. London: Hutchinson, pp. 156, 25s.



- K. R. HAMMOND and F. KERN (1960). *Teaching Comprehensive Medical Care : A Psychological Study of a Change in Medical Education*. Boston, Mass. : Harvard University Press, pp. xxii+642, 89s.,
- E. GLOVER (1960). *The Roots of Crime : Selected Papers on Psycho-Analysis*, Vol. 11. London : Imago, pp. xiii+422, 45s.
- J. G. CROWTHER (1960). *Francis Bacon : The First Statesman of Science*. London : Cresset Press, pp. xv+362, 35s.
- B. CROTHERS and R. S. PAINE (1959). *The Natural History of Cerebral Palsy*. Cambridge, Mass. : Harvard University Press ; London : Oxford University Press, pp. xi+299, 32s. 6d.
- R. G. ANDRY (1960). *Delinquency and Parental Pathology*. London : Methuen, pp. xv+173, 21s.

#### THE XIV INTERNATIONAL CONGRESS OF APPLIED PSYCHOLOGY, COPENHAGEN, AUGUST 13-19, 1961

The XIV International Congress of Applied Psychology will be held in Copenhagen at Christiansborg Castle, the Danish Parliament Building, August 13-19, 1961. Application forms and detailed information will be mailed during fall 1960 directly to registered psychologists in all countries.

Symposia are scheduled within most areas of applied psychology. A limited number of individual papers can be submitted in all fields. If you would like to present a paper, you must send a one page abstract to the Secretariat before January 1, 1961. Official languages : English, French, German. Simultaneous interpretation will be given.

Sightseeing and shopping tours will be arranged for the families of the delegates. Student baby sitters will take care of the youngsters.

President of the Congress : Professor E. Tranekjar Rasmussen, The Psychological Laboratory, University of Copenhagen.

Program Secretary : Assistant professor Gerhard Nielsen, The Psychological Laboratory, University of Copenhagen.

Secretariat : The XIV International Congress of Applied Psychology, 19, Sankt Pederstraede, Copenhagen K., Denmark.



## LEVELS OF ATTAINMENT OF EDUCABLE SUBNORMAL ADOLESCENTS

By R. E. MORÁN

*(Assistant Professor of Psychology Faculty of Education, University of Puerto Rico).*

**SUMMARY.** This study is an attempt to evaluate, by means of specially designed or adapted tests, the academic achievement of a representative sample of 300 educationally subnormal adolescents in the 'top' classes of London E.S.N. secondary schools. Their results are compared with those of 'normal' children with the same chronological ages in a secondary modern school and with those of children of similar mental ages in a junior school.

The over-all results of the E.S.N. adolescents show a wide range of variability, but generally fall below those of the junior school in every area tested, except craft. However, it was observed that in those schools where there were emphasis on and enthusiasm for a certain subject, in certain cases, this effected a relatively superior achievement in this subject, sometimes at the expense of others in the curriculum. It, therefore, seems that if the same enthusiasm and emphasis prevailed in all E.S.N. schools, and were directed at all subjects, combined with better teaching techniques and sound organization, an over-all rise in the general level of attainment could possibly be accomplished.

However, the writer suggests that the curriculum, although incorporating the academic and craft subjects, should be directed primarily towards vocational guidance and social maturity, rather than depending mainly on verbal facility, in which research has consistently shown the subnormal child to be most handicapped.

## I.—INTRODUCTION.

PSYCHOLOGISTS and educators have long been concerned with the standards of achievement to be expected of educationally subnormal children in view of their low mental capacities. Past studies in this field have yielded discrepant results. Thus, Merrill (1922), Wilson, F. T. (1926), and Lewis (1944) have shown that the academic achievement of subnormal children approximates and sometimes exceeds their mental ages. More recently, Bensberg (1953) has found that this population achieves essentially the same as the normative population for comparable mental ages.

On the other hand, Patterson and Phillio (1943) cite studies by Parker and Renshaw in which the educational achievement of sub-normal children was lower than their general intelligence. Studies in England by Burt (1922) and Sleight (1932) give similar results, and the former contends that, "... children who are retarded mentally appear still more retarded educationally." However, Gordon (1923), when studying subnormal children residing in the same district as many of those in Burt's research, observed that generally this group's attainment equates its mental ratio. Indeed, he found that his 'B' group (girls) had a spelling ratio (76.7) higher than its mental ratio (74.9).

Qualitatively, Wallin contends that no real moron (feeble-minded) can do third grade work (8-9 year level). Parker, Nemzek, Engel and others claim that subnormals finish school with attainments equivalent to fourth grade level (9-10 years) (Quoted from Merrill, 1922), while Burt (*op. cit.*), in England, has shown that educable defectives are usually capable of accomplishing work done in Standard II. (This would be comparable more or less to third grade level in

America, as children in England usually commence their formal studies a year earlier.)

However, regardless of whether the academic achievement of the subnormal child exceeds, equates or falls below his mental age, the procedures or instruments used in determining this have been constantly criticized by later research workers. For example, when the sub-normal's achievements are superior to his mental age, it is argued by some that this is due to a mere regression effect. Thus, if an arithmetic test correlates 0.70 with intelligence, then a child with an I.Q. of 60, M.A. 7 : 8, would be expected to have an arithmetic age of 9 : 3.

Conversely, when subnormals' achievements are below their mental ages, it is claimed by some that these children are penalized on attainment tests standardized on a nation-wide school population. The wide range in standards of teaching, school organization, equipment, material covered, etc., sets norms too high for the subnormal child. These tests are said to be designed as diagnostic tools to be used with normal and supernormal children rather than as research instruments for making comparative studies. There is also evidence suggesting that when mental age is determined by a test in which intelligence is measured by a single modality—reasoning, visual perception, verbal facility, etc. (e.g., Burt's study in which his Graded Reasoning Test was used)—the subnormal's achievement will tend to fall below his mental capacity. And, Patterson and Philleo (*op. cit.*) have pointed out that achievement norms based upon mixed groups at different I.Q. levels are misleading. Another valid criticism is that many investigators use timed group tests which also penalize the slow, apathetic subnormal child.

The present study is an attempt to investigate the level of academic attainments of subnormal children in relation to their mental capacities. The approach used attempts to obviate the criticisms of past studies. Only salient observations and results are reported here; for fuller details of interpretation and recommendations see Morán (1959). It is hoped that the results of this study will offer present and future educators and teachers a more lucid picture of the standard of scholastic work to be attained by children ascertained as educably subnormal and receiving education in special schools.

## II.—DESCRIPTION OF SAMPLE.

The vast majority of the children tested were from the London area. The experimental group included 300 subnormal adolescents, ages 14-16+ (mean C.A. 15 years 2 months), 160 boys and 140 girls, consisting of the 'top' classes of eight E.S.N. boys' senior schools (including one residential) and seven E.S.N. girls' senior schools (including one residential). Thus this experimental group is representative of 1,880 or about 12 per cent. of the total 2,180 London senior E.S.N. day school children. Their I.Qs. on the Terman-Merrill range from 41-90, mean 65.7 (boys' mean 68.5 and girls' mean 62.9) and on the Progressive Matrices\* from 55-99, mean 73.7 (boys' 77.7 and girls' 69.8). The mean Terman-Merrill mental age of the experimental group was 9 : 5 (boys 9 : 10 and girls 9 : 0); mean Progressive Matrices mental age 10 : 7 (boys 11 : 3 and girls 10 : 0).

The controls were drawn from one junior and one secondary modern school. Fifty children, ages 7-11 (mean C.A. 9 : 10, mean P.M. I.Q. 105, mean P.M.M.A. 10 : 4) were selected at random from among 458 pupils of a three-streamed mixed junior school, with the same proportion of boys to girls as in the experi-

\* The conversion table for expressing Progressive Matrices scores as M.As. and I.Qs. prepared by the Maudsley Hospital Institute of Psychiatry, was adopted (Gwynne Jones, 1956).

mental group. This junior group offers a sample of the attainments of normal children with similar mental ages to those in the experimental group.

To determine a level of academic achievement for adolescents of the same chronological age as those in the experimental group, approximately fifty boys and girls (mean C.A. 14:4, mean P.M.M.A. 12:3, mean P.M. I.Q. 88) were selected from a three streamed secondary modern school. Both the junior and the secondary modern schools draw their enrollment from catchment areas similar to those in which reside the majority of the E.S.N. children, viz.: working class and lower middle class residential districts.

### III.—TESTS ADMINISTERED.

All the tests used in the present study, with the exception of those of reading and intelligence, were specially adapted or constructed by the writer, and were administered as un-timed group tests in the following order:

*Mental Arithmetic*: An adaptation of *Burt's General Oral Test VIII*, in which a large proportion of the problems involved money transactions and were presented in meaningful life-like contexts.

*Craft* (Boys and Girls): The boys' and girls' craft tests are completely different in content, the former measuring, among other things, a specific knowledge of tools and their usage, while the latter is predominantly a test of general household knowledge and 'common sense.' Both tests are presented pictorially and were especially constructed by the writer for this study.

*Language*: This test, also specially constructed by the writer, is an attempt to measure the most elementary language usage and common grammatical errors committed by subnormal children.

*Spelling*: An adaptation of *Cattell's "Midland Tests 3"* was made by the writer so as to include words least affected by his non-English accent.

*Reading*: The *Vernon-Warden Group Reading Test A* was used, which is composed of two parts: word recognition and comprehension.

*General Information*: This test, constructed by the writer, is an attempt to obtain an indication of the subnormal's general knowledge, and contains sixty questions which cover the following areas: elementary Geography, History and General Science, Religion, Everyday Observation, Sports, Current Events Laws and Regulations, Civics and Government, Health and Hygiene.

*Intelligence*: *The Progressive Matrices* (1938, Revised Order 1956) was administered by the writer as an un-timed group test. *Terman-Merrill*: the most recent I.Q.s. of the experimental group on this test were also used as criteria for comparing attainment test scores.

Further included was an exercise in which the examinee was asked to write correctly and legibly his name, address, date of birth, the name of this school, and expected vocation. The writer also prepared a questionnaire which was used as a guide for eliciting Heads' and teachers' opinions and attitudes on the present E.S.N. curriculum, organization, etc.\*

### IV.—METHOD.

In order to assure a more uniform and perhaps more accurate measurement of the subnormal child's scholastic achievement, instead of depending solely on standardized tests, new tests were constructed and old ones specially adapted for this study. Test items, whenever possible, were based only on material which, according to the Heads and teachers, was universally taught in senior

\* A Social Maturity Scale, composed by the writer, and filled out by the teachers, was also included in the test battery, the results of which are found in Moran (*loc. cit.*)



E.S.N. schools. Teachers also offered suggestions as to how items should be worded in keeping with the subnormal child's paucity of vocabulary. Selecting items in this manner gives, to a great extent, a content validity to the tests used, in so far as the teachers would not approve the inclusion of items which they felt were not suitable for their children.

In order to mitigate the subnormal's sense of failure, items of tests specially constructed for this study, instead of being presented in order of difficulty, were arranged so that 'easy' and 'difficult' ones were interspersed throughout. Instead of using tests standardized on a nation-wide school population, the norms used here for comparison of the subnormals' academic achievement were calculated from the results of the average children with the same chronological age in a secondary modern school, and those of 'normal' children with similar mental ages in a junior school.

It is cogently argued that subnormal children are penalized on verbal loaded tests, and that I.Qs. obtained by tests measuring a single modality are not reliable measures of an individual's total mental capacity. To meet this criticism, in this study the mental capacity of the children of the experimental group is expressed according to their most recent I.Qs. on the Terman-Merrill, along with their results on the Progressive Matrices.

Tests were constructed so as to be given during one class period. A difficult test, e.g., arithmetic, was usually followed by one with more appeal, e.g., craft. Tests were not given during a rest period, activity, etc., and if possible neither on a Monday nor Friday.

The size of any one group was limited to twenty, and before commencing each sub-test, some time was dedicated to establishing 'rapport' and motivating the group. The sub-test papers were distributed separately; remaining tests were kept out of sight and 'retrieved' in such a manner as to create anticipation, thus arousing interest through curiosity.

Tests were constructed in such a manner as to require no reading ability, items being read aloud by the writer to the experimental group, which consisted largely of semi-literate or illiterate children.

The mean Mental Ages of the 8-9 and 10-11 year groups, normal juniors (calculated from their converted Progressive Matrices I.Qs.) with their respective mean scores on each attainment test, were used as a criterion from which were extracted approximate attainment ages for the different tests. This procedure, which has been used by Schonell and Schonell (1950), is based on the assumption that the test scores of the junior group would be indicative of the work done by the average child from a working class environment, at the chronological age level of 8-11, which is, in effect, the mental age range of the experimental group. The various test scores expressed in this way represent a rough educational age standardization and so make possible comparisons of each school on each test. It also provides an appreciation of the overall level of attainment obtained by the experimental group. Scores expressed in this manner as attainment ages have obvious limitations; though, as Burt (1922) pointed out, "... the mental year, however crude, would seem as a unit, to be sufficiently exact, as it is eminently serviceable." Boys' and girls' test scores were tabulated separately.

In order to investigate the dispersion of scores, and to appreciate the percentage of children in the experimental group who are working up to the capacity of 'normal' children of the same chronological age in the ordinary school, the median and the 10th percentile were calculated from the raw scores of the controls in the secondary modern school on each attainment test.



## IV.—RESULTS.

The rank correlation between the mean P.M. I.Qs. and the T.M. I.Qs. of the experimental group schools was 0.73, which is reasonable although lower than the figure given by Raven, the author of the Progressive Matrices test. This may be due to three factors: homogeneity of the present group, the low 'floor' of the Matrices which seems to favour the sub-normal child, and the fact that the tests measure different mental processes.

TABLE 1  
ATTAINMENT TEST RESULTS OF EXPERIMENTAL GROUP (MEAN AGES).

	Boys	Girls	Total Exp. Group
Mean C.A. ....	15:3	15:1	15:2
Mean PMMA ....	11:3	10:0	10:7
Mean TMMA.....	9:10	9:0	9:5
Reading .....	8:3	8:3	8:3
Arithmetic.....	8:5	6:7	7:6
General Information	9:6	7:11	8:8
Language .....	6:10	7:0	6:11
Spelling .....	7:3	7:8	7:5
Craft .....	11:8	10:8	11:2

TABLE 2  
PERCENTAGES ABOVE, BETWEEN AND BELOW THE INDICATED MODERN SCHOOL PERCENTILES.

Boys			
Subject	Above 50th P.	Between 10th and 50th P.	Below 10th P.
Language .....	8.3	28.0	63.7
Reading .....	0.7	12.4	86.9
Spelling .....	3.8	23.8	72.4
Arithmetic.....	2.4	48.2	49.4
General Information	4.8	44.4	50.8
Craft .....	37.8	39.5	22.7
Soc. Mat. ....	42.9	50.0	6.7
Girls			
Language .....	10.3	26.4	63.3
Reading .....	1.7	12.1	86.2
Spelling .....	7.0	25.7	67.3
Arithmetic.....	0.0	19.8	80.2
General Information	0.4	26.4	73.2
Craft .....	9.4	48.2	42.4
Soc. Mat. ....	39.4	57.6	2.8

The results of the experimental group on the attainment tests expressed in mean ages are shown in Table 1, and in Table 2 their results, as compared with the controls in the secondary modern school are indicated by percentiles. As

TABLE 3

E.S.N. SCHOOLS RANKED ACCORDING TO MEAN P.M.I.Qs. WITH CORRESPONDING RANKS OF T.M. AND ATTAINMENT TESTS.

Boys' Schools	PMIQ	TMIQ	PMMA	TMMA	R	A	Sp	Lang	G.I.	Cr.
AB	1	7	1	5	9	4.5	13	11	4	6
BB	2	6	2	7.5	13.5	3	14	14	10	7.5
CB	3	3.5	4	4	15	6	15	11	7	5
DB	4	3.5	3	2	1	8	1	1	3	7.5
EB	5	1	5	1	2	1	4	3.5	1	4
FB	6	8	8	9	7	7	7.5	11	2	2
GB	7	2	9	3	12	4.5	10.5	5.5	5.5	3
HB	8	9.5	6.5	7.5	3	2	10.5	10	5.5	1
Girls' Schools										
1G	9	13	6.5	12	10.5	10	5.5	7	8	7
2G	10	5	11	6	3	12	3	2	12	6
3G	11	11	10	10	8	13	7.5	3.5	14	3
4G	12	9.5	12	11	6	11	12	7	9	1
5G	13	12	13	13	10.5	9	5.5	7	11	4
6G	14	15	15	15	13.5	15	9	15	15	2
7G	15	14	14	14	3	14	2	5.5	13	5

seen in Table 3 the boys' schools consistently rank higher than the girls' in both intelligence tests, with one exception on the Terman-Merrill. However, both sex groups tend to score higher on the Progressive Matrices than on the Terman-Merrill. These findings are contrary to those of Walton (1955) whose results obtained from a group of subnormal children, showed differences in I.Qs., but mostly in the direction of lower Matrices results.

There exists a wide range of variability in the level of attainment among the schools of the experimental group. However, according to analyses of variance, significant differences between schools occurred only in Arithmetic, Reading and General Information. Such variations, it seems, may be attributed to differences in teaching techniques, school organization, and emphasis on or enthusiasm for specific subjects in the various schools.

As shown in Table 3, there exists little or no relationship between the ranking of the schools of the experimental group on either test of intelligence and their ranking on the various attainment tests. However, schools which occupied the highest positions in intelligence or in the language subjects, tend to score low in craft: and the boys' school ranking highest in craft scored lowest in intelligence.

It is generally believed that girls excel boys in linguistic skills, and that boys are superior to girls in arithmetic and general knowledge. This has also proved true in the present study. Thus, we find in the language test the mean score of the girls of the experimental group was higher than that of the boys, whose mean language score was lower than in any other subject tested. However both sexes scored considerably lower than the junior group. This is shown in Table 1.

In reading we find that, according to the results on the tests used in the present study, the mean score of the boys and that of the girls of the experimental group were identical, which suggests a slight superiority in favour of the

girls who tend to have lower I.Qs. This superiority of the girls in reading is in keeping with the findings of most studies. However, the mean reading score of the experimental group was significantly lower than that of the controls with the same mental ages. It is interesting to note that the reading test used was standardized on a primary school population and that the time limit of 20 minutes for the age group 6-9 was permitted for the experimental group. If mental age is used as a standard, there seems to exist a wide discrepancy between reading attainment and mental capacity in the experimental group. This is in contrast to the findings of Wilson, F. T. (1926), Bensberg (1953) and MacIntyre (1937). As seen in Table 3, there exists little relation between the rank orders in reading and intelligence in the experimental group.

The results of the experimental group in spelling in the present study are more in keeping with those of Burt (1922), and Sleight (1932), who found the subnormal child exceptionally inefficient, than with those of Merrill (1922) Wilson (1926) and Gordon (1923) who give evidence of these children's relative efficiency as compared to their mental ages. This can be seen in Table 1.

The boys of the experimental group were significantly superior to the girls in arithmetic, a fact generally accepted if not satisfactorily explained. However, the total mean arithmetic score of the experimental group is lower than that of the junior group. It must be remembered that the items of the arithmetic test were adapted so that most problems involved money transactions, which is in accordance with these children's day-to-day experiences. Thus, the experimental group's results in arithmetic in this study are contrary to those found in others (Wilson, *op. cit.*, Merrill, *op. cit.*, Bensberg, *op. cit.*, Patterson and Philleo, *op. cit.*), in which subnormals' scores in arithmetic equate or are above their mental ages.

The boys' score in the experimental group on general information is markedly superior to that of the girls. This superiority of boys over girls in general information at all levels of intelligence is in keeping with the findings of other studies (Terman, 1926, Inman, 1939). However, as is shown in Table 1, the children of the junior school obtained a higher mean general information age than the adolescents of the experimental group whose average total life experience is 5 years 4 months greater. Nevertheless, many children of the experimental group showed a relatively wide range of general knowledge, which, in many instances exceeded that expected of them by their teachers. Further enquiry by the writer revealed that many of these children are capable of recalling and associating facts when these are presented through visuo-auditory media such as T.V. and newsreels.

The boys' and girls' craft tests were completely different in content, and therefore, the scores of the two sex groups were tabulated separately. The mean craft age of the boys of the experimental group is higher than that of the junior school and approximates the mean score of the secondary modern school. In fact, one school of the experimental group, the one with the lowest mean I.Q. on both intelligence tests, exceeds the mean score of the secondary modern school. The mean craft age of the girls in the experimental group is also higher than that of the girls in the junior school. However, the E.S.N. girls' mean craft age is significantly below that of the girls in the secondary modern school. The mean ages of both sex groups is higher in craft than in any other subject tested. However, a scrutiny of individual school results reveals that, in craft, E.S.N. schools tend to rank in almost direct inverse order to both their P.M. and T.M. mental age rankings. This can be seen in Table 3.

Table 4 shows the results (per cent. correct) of the exercise in which the examinee's sole task was to write legibly and correctly his name (Christian and

surname), date of birth (day, month, year), present age, age next birthday, his present address, and the name of his schools. As can be seen, the results of the children of the junior school in the age group 10-11 are comparable to those of the A and B streams of the secondary modern school, and are significantly higher than those of the adolescents in the experimental group. There is a slight sex difference in favour of the girls of the experimental group in all parts of this exercise.

TABLE 4

RESULTS OF JUNIOR, SECONDARY MODERN AND E.S.N. SCHOOLS ON QUESTIONNAIRE.

School	Group or Sex	Name	Date of Birth	Present Age	Present Address	Name of School
JS	10-11	100.0	92.0	96.0	92.0	92.0
JS	8-9	72.0	64.0	80.0	52.0	56.0
JS	Total	86.0	78.0	88.0	72.0	74.0
SMS	'AB'	100.0	96.0	100.0	100.0	92.0
SMS	'C'	95.4	50.0	90.9	81.8	59.0
SMS	Total	97.7	73.0	95.4	90.9	75.5
ESN	Boys	71.9	61.4	67.8	70.1	67.8
ESN	Girls	78.8	66.9	75.0	77.9	70.5
ESN	Total	75.3	64.1	71.4	74.0	69.1

Table 5 gives an overall comparison of the mean results of the experimental group on the spelling, reading and arithmetic tests used in this study with the results obtained by the E.S.N. schools themselves, using Burt's or Schonell's standardized attainment tests, together with the results cited earlier by Burt. It will be seen that the findings are generally quite similar.

TABLE 5

COMPARISON OF VARIOUS ASSESSMENTS OF ABILITIES OF E.S.N. PUPILS (MEAN AGES).

Subject	Burt's Study (1919) <sup>1</sup>	Moran (1958)	School Assessment <sup>2</sup>
Spelling .....	6 : 5	7 : 5	7 : 3
Reading .....	6 : 8	8 : 3	7 : 9
Arithmetic .....	7 : 2	7 : 6	7 : 10
Handwork .....	9 : 8	11 : 2	—
Average .....	7 : 5	8 : 7	7 : 7
Average M.A. ....	8 : 1 (Binet)	9 : 5 (Terman-Merrill)	
Average C.A. ....	14	10 : 7 (Progressive Matrices)	
Average I.Q. ....	62 (Binet)	15 : 2	
		65.7 : (Terman-Merrill)	
		73.7 (Progressive Matrices)	

<sup>1</sup> See Burt (1922, p. 337).

<sup>2</sup> Mean Ages calculated from results of most recent attainment tests given in E.S.N. schools (Burt or Schonell).



## V.—DISCUSSION.

The results of the experimental group, as compared with those of the junior and secondary modern school, strongly suggest that a small percentage of adolescents of limited intelligence are capable of and are actually performing at a level of scholastic achievement comparable to that of children with similar mental ages in a junior school or in a 'C' stream of a secondary modern school, as claimed by Parker, *et al* (*op. cit.*); that there exists an undetermined number of children who are potentially capable of working at a higher level, but who may not be doing so, due to factors other than that of intelligence; and that, under the present administrative system and school curriculum, the majority of the subnormal adolescents approaching school leaving age tend to score lower, on the attainment tests used (with the exception of craft) than children in the junior school with similar mental ages, and residing in a similar socio-economic environment. In other words, their average level of scholastic achievement is equivalent to that of a second form in the ordinary junior school, as was originally claimed by Burt (1922).

However, it must be emphasized that the mental ages of E.S.N. children should be considered only as an indicator, and not as a limit of academic potential. Moreover, the type of test used makes a considerable difference, e.g., individual, group, verbal, non-verbal, etc., For it was observed in the present study that an M.A. calculated from a verbally loaded test will generally be lower than if calculated from a test with a high 'g,' 'k' or perceptual loading; and that the subnormal child's attainments tend to approximate more closely to this verbal M.A. On non-verbal tests, subnormals tend to score higher, thus increasing the discrepancy between mental age and academic achievement. This might explain, in part, the discrepancy between the findings of the present study and those of other similar investigations, in which, in most instances verbally-loaded tests were used, and in which the attainment of subnormals equated or surpassed their M.As. It is observed, however, that in the present study, the results of the experimental group are more in keeping with those of English investigators than with those of American. Thus, one may hazard the hypothesis that the closer approximation of subnormal children's attainments to their mental capacity in American studies than in English ones may be due to a difference in the standard of work at the various age or grade levels in the two countries. Some evidence supporting this hypothesis is presented in a comparative study by Pidgeon (1958). However, the subnormal's relative ability to deal with tasks involving spatial or perceptual rather than verbal factors, may indicate a need for re-organizing the present E.S.N. curriculum, and re-examining the present philosophy and practice of education, in which emphasis is still placed on those disciplines in which the subnormal's scholastic success depends largely on his competence in handling verbal material. Thus, it seems logical that the curriculum of secondary special schools or classes for subnormal children should be based on those mental processes in which experimentation has shown these children to be relatively more competent, i.e., according to their mental assets rather than to their weaknesses. A lead in this direction was taken by Duncan (1942) in England and Winterbourn, R., (1944) in New Zealand, who have claimed favourable results. Unfortunately, these claims have not been substantiated by other workers. Nevertheless, until these mental functions have been scientifically determined, the E.S.N. school curriculum could be planned around a programme, incorporating the necessary basic and craft subjects, but mainly directed towards inculcating attitudes and habits necessary for vocational and social competency. Such a programme would be reinforced by continual and varied experiences in the area covered in the

general information test. Television could be used as a supplementary medium of instruction by which the subnormal child could be provided with information in a form in keeping with his relatively superior visuo-perceptual ability.

It is also suggested that, on devising tests for E.S.N. children, the accepted practice in psychometry of presenting items in order of difficulty is not advisable. Rather, difficult and easy items should be interspersed, commencing and terminating with the easier ones.

ACKNOWLEDGMENTS.—The writer wishes to acknowledge his indebtedness to the following: The Department of Education of Puerto Rico for providing the opportunity for this study; London County Council for granting permission to visit and administer tests in London schools; all the staff of the E.S.N. schools visited for their whole-hearted co-operation; the Central Research Fund of the University of London which helped defray the cost of test material; and Professor P. E. Vernon, University of London, under whose supervision and guidance this study was carried out.

## VI.—BIBLIOGRAPHY.

- BENSBERG, J. G. (1953). The relation of academic achievement of mental age, institutionalization and etiology. *Amer. J. Ment. Defic.*, 58, 327-330.
- BURT, C. (1922). *Mental and Scholastic Tests*. London: P. S. King.
- DUNCAN, J. (1942). *The Education of the Ordinary Child*. London: Nelson.
- GORDON, H. (1923). *Mental and Scholastic Tests Among Retarded Children*. Educ. Pamphlets, No. 44, H.M.S.O.
- GWYNNE JONES, H. (1956). Comments on: The validity and inter-changeability of the Terman-Merrill and matrices test data. *Brit. J. Educ. Psychol.*, 26, 141.
- INMAN, R. D. M. (1939). General knowledge and intelligence. *Brit. J. Educ. Psychol.*, 9, 72-86.
- LEWIS, W. D. (1944). The relative intellectual achievement of mentally gifted and retarded children. *J. Exp. Educ.*, 13, 98-109.
- MACINTYRE, E. M. (1937). Teaching reading to mentally defective children. *Proc. and Addresses of 61st Annual Session of Amer. Assoc. on Ment. Defic.*, 42, No. 2, 59-67.
- MERRILL, M. A. (1922, 1924). On the relation of intelligence to achievement in the case of mentally retarded children. *Compar. Psychol., Monogr.*, 2, Serial No. 10.
- MORAN, R. E. (1959). *An Empirical Study of a Representative Group of Educationally Subnormal Adolescents*. Unpublished Ph.D. Thesis, Univ. of London.
- PATTERSON, R., and PHILLEO, C. (1943). Academic achievement of differentiated groups at the Moron level. *Amer. J. Ment. Defic.*, 48.
- PIDGEON, D. A. (1958). A comparative study of basic attainments. *Educ. Research.*, 1, No. 1, 50-68.
- SCHONELL, F. J. and SCHONELL, F. E. (1950). *Diagnostic and Attainment Testing*. Edinburgh: Oliver and Boyd.
- SLEIGHT, G. F. (1932). *The Diagnosis and Treatment of the Dull and Backward Child*. Unpublished Ph.D. Thesis, Univ. of London.
- TERMAN, L. M. (1926). *Genetic Studies of Genius*. Stanford Univ. Press.
- WALTON, D. (1955). The validity and interchangeability of Terman-Merrill and matrices test data. *Brit. J. Educ. Psychol.*, 25, 190-194.
- WILSON, F. T. (1926). Some achievements of pupils of the same mental age but different I.Qs. *J. Educ. Research*, 14, 43-53.
- WINTERBOURN, R. (1944). *Educating Backward Children in New Zealand*. New Zealand: Whitcombe and Tombs.

(Manuscript received, November 28th, 1959.)

# OBSERVATIONS ON RETEST DISCREPANCY IN MENTALLY SUB-NORMAL CHILDREN

By D. H. STOTT

*(Psychology Department, University of Glasgow.)*

**SUMMARY.** Retest scores of subnormal children were examined in order to check the reliability of previous retests and to estimate to what extent the discrepancies might reflect real changes in the rate of mental development as well as errors of standardization.

The discrepancies were considerably greater than those obtained from the standardization of the Binet by Terman and Merrill. The usual tendency towards increasing discrepancy with time-interval was confirmed, except for an apparent practice-effect at short intervals. The gains and losses were much more heterogeneous than could be accounted for by intra-test error. Boys lost significantly less than girls.

It was concluded that the retest discrepancies might in large part be due to real variations in the rate of mental development (including such as are dependent on changes in motivation).

## I.—SCOPE AND PROBLEMS OF THE ENQUIRY.

It was observed in a limited number of special schools for educationally sub-normal children (not included in the survey reported below) that I.Q.s. on retest often showed greater fluctuation than that considered the rule for children of this type (Terman and Merrill reporting an average difference between scores on the L and M forms for the below-70's of only  $2\frac{1}{2}$  points). The present enquiry is an attempt to assess the generality of this observation and the extent to which the discrepancies on retest may reflect real changes of cognitive or conative development.

It is first necessary to consider causes of discrepancy which arise from the procedure of testing or from the nature of the tests. The chief of these are probably (a) the varying ability of different testers to secure rapport; (b) the use of different tests; (c) test-sophistication; and (d) faulty test-standardization, especially in respect of subnormal children of school age.

Each of the above factors is examined below, but the fourth involves some theoretical points better mentioned at this stage. Discussing the deviations from variance-constancy found in the standardization of the Terman-Merrill revision of the Stanford-Binet, upon which their Correction Table is based, Fraser Roberts and Mellone (1952) write: "It is most unlikely that the variations in variance have any particular meaning in terms of real mental growth; they are, doubtless, due to the way the scale is constructed and the way the measurements are used." They give no reason for this conclusion, although they quote McNemar (1942) as pointing out that the 6-year level of the test has insufficient items of medium difficulty, while the 12-year level has too many.

It cannot, however, be taken for granted, as do Roberts and Mellone, that all observed variations of variance as between one-year-level and another are due to faulty standardization. They may also be caused, wholly or partly, by real variations in mental growth as between sub- and supra-normals. The latter hypothesis is the more recent one, and has been rendered feasible by the general recognition of large environmental influences in mental development and of the contemporary nature of mental measurement (Vernon, 1959). It is represented by Clarke and Clarke (1958), who take up what must be considered



*prima facie* a biologically reasonable position: "During childhood and adolescence, mental (like physical) growth does not proceed at a uniform rate and the individual's position with respect to others in his group tends to vary from year to year, this being reflected in changes in I.Q." In support of this view, they refer to an estimate, based on Thorndike's summary mentioned below, that correlation between test and retest decreases on average by 0.04 per year. Since, they argue, errors of measurement are unlikely to be cumulative, "this must be a reflection of real change in intellectual level due to different rates of development," (their italics). However, they do not refer to the alternative possibility that discrepancies widening over certain age-ranges may arise from faulty test standardisation. They quote certain studies which report large changes in I.Q. Of these, that of Honzik, MacFarlane and Allan (1948) is the most spectacular; but, as the Clarkes point out, the I.Q. differences represent the greatest divergences registered by each individual between the ages of 6 and 18 years, which not only maximises accidents of the test-situation, but eliminates a number of lesser variations and identical retest scores which would give a more balanced picture. Vernon also points out (1955) that the children who were the subject of this research were of superior ability, and the Clarkes agree that this would give a great range for variation over a large number of years. The Clarkes' own study (1953) is also relevant: they found, in the case of over 100 adult mental defectives, that those who came into a mental institution from very bad home environments showed a mean rise of 10 I.Q. points two years after admission, compared with those from less adverse environments, who only showed the equivalent of the normal effects of test-practice.

Thorndike (1933) made a rough average of a number of studies which had found increasing discrepancy with length of time between tests: whereas the correlation for an immediate retest fell close to .90, for a test after a 5-year interval it was estimated to be about .70, giving an average annual decline of .04. In a later review of work on I.Q. constancy, Thorndike (1940) quoted a study by Brown (1933), which reported a drop of retest correlation from .86 after an interval of less than two years to .61 after five years or more; and a study by Hirsch, reanalysed by Miller (1933), which reported a progressive increase in the median difference in I.Q. from 5.4 at one year interval to 7.3 points at five-years. It seems, in summary, well established that divergences on retest tend to be greater the longer the interval between the tests.

The present enquiry is based on the assumption that if retest discrepancies reflect an element of change in the rate of mental growth, they will show two characteristics. First, they will be heterogeneous; that is to say, even though there may be a net trend upwards or downwards, there will be a sufficient number in the opposite direction to suggest significant individual variation, and of those in the direction of the net trend the variance will be large. Second, the tendency for the discrepancies to be greater the longer the interval between testings should ideally also apply to those against the net trend.

## II.—DESIGN OF THE EXPERIMENT.

By arrangement with the Special Schools Association, head-teachers of day special schools for educationally sub-normal children (of I.Q. largely within the range 50-75) over the greater part of England were asked to supply data in respect of the most recent fifty children, or any smaller number, who had been tested twice or more times. In some areas the consent of the local education authority was not forthcoming, and some schools had no such data; however, the thirty-four who supplied utilisable material probably represented a fairly random sample. It was emphasised to the head-teachers that the test results



should be continuous series, and not examples of divergence or otherwise. The *pro forma* asked for the ages of the children at each testing, the test used, and who tested.

Data were obtained for 952 children. Some of these, however, had been tested three, four or five times. In these cases the logical procedure was to compare each test-result with each other, so that three testings yielded three comparisons, four six and five ten. The total of comparisons was reached 1358. Of these, both test and retest were made by the same tester in 830 cases, and by different testers in 528. All were the appointed examiners of local education authorities, namely educational psychologists or medical officers. Classification by length of interval between tests had to be arbitrary, but it was thought best to make sure of including within the shortest interval what were intended to be annual retests. Consequently, to allow for the possibility of these being made during the fourth school term, that is, a term in arrear, the shortest interval was fixed at 1 year 4 months. The remainder were fixed as follows: 1 year 5 months to 2 years 4 months; 2 years 5 months to 3 years 4 months; and 3 years 5 months or more. No trial was made of any alternative division, so that the above may or may not approximate to the one which would bring out the greatest interval-differences.

### III.—FINDINGS.

The distribution of the total retest discrepancies is shown in Figure I. 28.0 per cent. were of 10 I.Q. points or more; 11.1 per cent. of 15 points or more; and 3.5 per cent. of 20 or more. In the table which Vernon (1957) publishes in *Secondary School Selection* (Appendix B), for 29 per cent. of I.Q.s. to change 10 points or more, and 3 per cent. to change 20 or more corresponds to a correlation of .80. Terman and Merrill claimed a reliability of .98 for the below-70's and .95 for the 70-89 group. There are, however, several possible explanations of this disparity. The examiners for the T-M standardization were

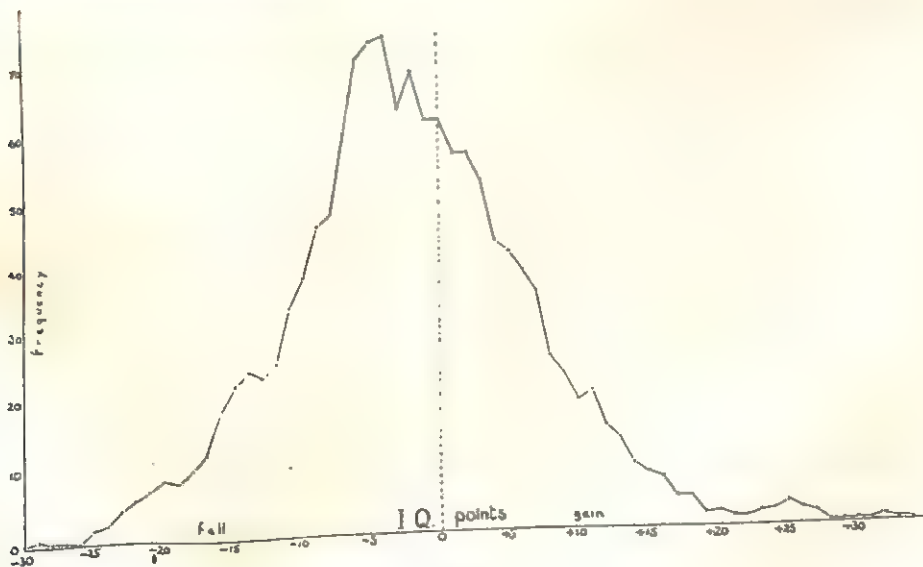


Fig. I.

working self-consciously upon a project, whereas the testing in the present sample was done in the ordinary course of administrative ascertainment and supervision. The intervals between retests were shorter in the standardization, so that there would have been a greater tendency for a mean rise from practice-effect to cancel deterioration of performance by subnormals. Lastly, it can be calculated from the distribution graph (Terman and Merrill, p. 37) that their sample covered only about 144 children of all ages of I.Q. below 75, that is, only about one-tenth of the present sample.

The mean discrepancy, disregarding direction, was 7.098 points. It is seen that there is a consistent displacement to the 'minus' side (falls in I.Q.). If the fifty retests which registered no change are allocated equally between the rises and falls, the mean of the rises is 6.512 and the mean of the falls 7.449; but since there were nearly twice as many falls as rises the net mean discrepancy, taking account of direction, was -2.194. Despite this general tendency, consistent with the findings of Roberts and Mellone, for the I.Qs. of subnormal children of school-age to fall, there is a sufficient heterogeneity of retest discrepancy to render feasible the hypothesis that the mental capacity of a significant proportion of the children changed significantly. For the present argument a discrepancy of plus or minus 10 points may be taken as an arbitrary criterion of significant change (for whatever reason). Assuming that some systematic correction is needed on account of faulty test-standardization, the above will have to be adjusted to allow for the net mean fall of just over 2 points, giving as rough criteria of significant change gains of 8 and falls of 12 points. On this basis there were 166 significant rises and 178 falls, which combined amounted to 25.3 per cent. of the test-retest comparisons. Reference to the Roberts-Mellone Correction Table shows that for subnormal children the only gains needing correction occur in the age range of 5:0—7:11, and the corrections recommended are deductions of 1 to 4 points from 5:0—6:11 and 0 to 3 for 7:—7:11. In the present study 16.3 per cent. of the first tests in the total of retest comparisons were made when the children were in the former of these age-groups and 16.4 per cent. when they were in the latter. Thus, while the falls could be partly accounted for by errors of standardization, the gains could not be so, since either they ran counter to the recommended corrections or if in the same direction were not of the same degree. The heterogeneity cannot, therefore, in the main be due to faulty standardization, whatever this may mean.

The other feasible causes of retest discrepancy must be evaluated at this stage. The test used was almost exclusively the Terman-Merrill revision of the Stanford-Binet (the few others being one each on the Porteus and Alexander Performance and five unknown), so that the use of different tests as a reason for the discrepancies can be ruled out. Where, however, the children were tested by a different tester there is a tendency to greater variability. The mean discrepancy with the same tester was 6.80, and with different one 7.49. This is significant at between the 2 per cent. and 5 per cent. risk of chance ( $t=2.243$ ).

In view of the inferences made below from the tendency of the discrepancies to increase with length of retest interval, the possibility must be considered that the latter might in part be due to the greater probability of a different tester when the intervals were long. In fact, as Table 1 shows, the greater over-all mean discrepancy for 'different tester' arose entirely within the *shorter* intervals (up to 2:4 years). At the longer intervals it is seen that the means for 'different tester' were actually slightly less than for 'same tester.'

This disposes of the possibility that the change of tester might have produced the widening discrepancies over the longer intervals. It can also be said that while the human factor of the tester has some influence it is less than might have

TABLE 1  
MEAN RETEST DISCREPANCY (I.Q. POINTS).

	Same Tester	Different Tester
Retest Interval : 0—1 : 4 years .....	5.495	6.766
1 : 5—2 : 4 .....	5.530	7.628
2 : 5—3 : 4 .....	6.704	6.631
3 : 5 or more .....	8.916	8.598

been thought, and certainly not the main reason for the retest discrepancies observed.

The above-mentioned tendency for the test results to be more discrepant over the shorter intervals only with different testers seems an anomalous one. It may possibly be explained by test-sophistication. Elwood (1952) found that retests in which the same form of the Terman-Merrill scale were used gave slightly higher I.Q.s. than those with the other form. It is also possible that the mean rise in I.Q. which Rudolf (1949), using the Wechsler scales, found in institutionalised defectives after intervals varying from six to eighteen months was also due to test-sophistication.

In the administrative situation in which the present testings were carried out, the same tester retesting a child after a comparatively short interval would normally take the precaution of using the alternative form; another tester would probably have no record of which one was used for the earlier test. Thus, it is possible to hypothesise a practice-effect, or more accurately, a memory-effect, when subnormal children are confronted with the identical version of the Terman-Merrill up to intervals of 2 years or so. If this is well-founded, the discrepancies in question would be predominantly increases, despite the overall tendency to decrease. Table 2 shows that there is, indeed, a tendency in this direction: the only cells in which the plus discrepancies have higher means than the minus of the same retest interval are in the shortest interval for 'same tester' and the two shortest for 'different tester.' Nevertheless, this

TABLE 2  
MEAN PLUS AND MINUS DISCREPANCIES BY TESTER AND INTERVAL.

Retest interval :	Same Tester		Different Tester	
	-	+	-	+
0—1 : 4 years .....	5.154	5.896	6.500	6.936
1 : 5—2 : 4 years ....	5.591	5.545	6.928	8.660
2 : 5—3 : 4 years ....	7.162	5.763	7.007	6.279
3 : 5 years or over ....	9.517	6.831	9.103	6.884

defiance of the fall-tendency is not quite strong enough to be significant in terms of mean differences. In the different-tester group for the two shorter intervals, the mean of the + 's is 7.698 and of the - 's 6.750, giving a  $t$  of 1.467. However, if the number of cases in which the I.Q. rose is compared with the number of falls, irrespective of amount of either, the preponderance of rises within the two shorter intervals of the different-tester group is significant (Table 3) at the 0.1 per cent. level.

TABLE 3

NUMBER OF RISES AND FALLS IN I.Q. IN DIFFERENT-TESTER GROUP.

	Falls	Rises	
Two shorter intervals .....	96 (116.8)	113 (92.2)	209
Two longer intervals .....	199 (178.2)	120 (140.8)	319
	295	233	528

Expected incidence in brackets  
 $\chi^2 = 13.89$   $p < .001$

In general, therefore, the hypothesis of a degree of test-sophistication would seem to be borne out. On the other hand, since the tendency for the plus discrepancies to be greater with the shorter intervals of the same-tester group is so slight, this factor of test-sophistication may be discounted in the further treatment of the data below.

The tendency for the retest discrepancies to increase with test interval can be best studied, therefore, within the same-tester group of scores, and there are sufficient cases (830) for the different-tester group to be dispensed with.

In Table 4 the mean discrepancies within every interval are compared with every other. It is seen that there is a significant enlargement in the retest discrepancy as between every shorter and every longer interval, except for the shortest with the next shortest, and even this only just falls short of significance. And, as previous work has shown, the longer the retest interval the greater the discrepancy.

Such a trend is consistent with the hypothesis that subnormal children are subject to varying rates of mental growth. It is doubtful if such consistency could be obtained from incorrect test-standardization. Even if it were, the whereas in the present sample there is a slight tendency thereto in the rises besides the marked one in the falls (Table 2).

A further means of testing the Clarke's hypothesis of individually differing rates of development suggested itself during the treatment of the data. The present writer had observed in casework among backward children of school age puberty onwards. This was most marked where the mental retardation seemed to be due in part, or possibly wholly, to the type of congenital impairment which he had named Unforthcomingness (Stott, 1959, a, b). This is recognized by German and Austrian clinicians under the name of *Antriebsmangel* (lack of motivation), and is thought to be due to damage to the frontal lobes. The basic



TABLE 4

INCREASE IN RETEST DISCREPANCY WITH TEST-INTERVAL (SAME TESTER).

Test interval	1:5—2:4	2:5—3:4	3:5 or more
0—1:4	(a) 0.691 (b) 0.423 t=1.634 Not sig.	(a) 1.309 (b) 0.587 t=2.229 p<.05	(a) 3.421 (b) 0.552 t=6.197 p<.001
1:5—2:4		(a) 1.174 (b) 0.523 t=2.245 p<.05	(a) 3.386 (b) 0.483 t=7.011 p<.001
2:5—3:4			(a) 2.212 (b) 0.574 t=3.854 p<.001

(a) Difference in mean discrepancy.

(b) Standard error of same.

characteristic is a lack of the assertiveness which prompts a normal child to face strange situations and tasks, so that the unforthcoming child appears as a 'mousy,' confidence-lacking type (Stott, 1956, 1958). It is notable that boys of this type, but not in the writer's limited experience girls, tend during adolescence to acquire the assertiveness and confidence needed for tackling reading: and it is reasonable to suppose that they would also be able to face the tasks set by an intelligence-test with greater confidence and hence with a greater chance of success.

There thus emerged the possibility of a sex-difference in rate of mental development. To test this it was necessary to approach once again the heads of the thirty-four schools who had supplied the original data in order to find out the sexes of the children. This information was obtained for all but four schools, in which there had been a change of head or the children could not be identified. In one of these schools the sex of two children—boys with I.Q. losses of 12 and 15—was identifiable, and these were included, since they told against the hypothesis of sex-difference. The numbers of test-retest comparisons in respect of children for whom the sexes were known were 632 for boys and 643 for girls, giving a total of 1,276 compared with the original one of 1,358. The verification of a sex-difference required here, not the ascertainment of larger or smaller discrepancies as between boys and girls, but an over-all tendency for boys to gain more frequently, or lose less so, than girls. It was found that the boys suffered a net loss of only 1.189, compared with 2.985 for the girls. With  $t=3.719$ , this was significant at less than 1 in a 1,000 risk of chance. The sex-difference thus established must be held to be evidence of differing individual rates of mental development, for it is hardly conceivable that this could be due to intra-test factors. A corollary is that, if there remains any logical justification for a correction table, there should be different ones for each sex.

## IV.—CONCLUSIONS.

The tendencies for retest discrepancies to be heterogeneous, to enlarge with the length of the interval between tests, and for backward boys to lose significantly less on retest than backward girls, suggest that the observed variations in variance of the Binet tests are due in large part to real changes in the rate of mental development of subnormal children of school-age. This is not to deny that part of the variation is also due to faulty standardisation. Close study of the percentages of children of various ages passing each age-level of the test, as given in McNemar's (1942) Table 26, reveal obvious inconsistencies of standardization, and it would be surprising if these were not reflected in distortions of variance. For example, the 5-year level is much too easy for children of 5 years and over—presumably also for subnormals for whom it is liminal—while it is much too hard for children under 5 years. The 9-year level is again too hard for children below that age. Nevertheless, the utility of a correction-table is doubtful, in view of the difficulty of isolating the influence of faulty standardization from real accelerations or decelerations of mental development. In using such a table there is a danger that real changes of mental growth may be concealed.

It may also be asked whether faulty standardisation has any meaning except in the sense of a faulty methodology, such as reliance upon too small samples of subjects. Cannot the curious trend of the mean percentages of normal children of different ages who pass the 5-year level, referred to above, be due to a jump in the intellectual development of children between 4 and 5 years? It may, indeed, be virtually impossible to construct any battery of test-items which did not show this irregularity. Changes of variability, including those of standard deviation, do not, then, pre-suppose faulty standardization. The concept of standardization depends upon the assumption that variance will not change for natural reasons. But in so far as the contemporary mental ability of any individual is the end-product of a cumulative development and interaction, one would expect a widening of the gaps between the well and not-so-well mentally equipped, and between the most and least environmentally favoured. If we ignore this, and 'standardize' a test so as to make the range of 'intelligence' similar at all ages, we may cover up this or other natural changes in variability. Our test would in effect be neatly, but falsely, standardized by giving progressive bonuses to the weak or the environmentally less favoured. In any case it becomes necessary to adopt a more empirical approach to standardization.

Finally, it cannot be assumed that such changes of mental development, so called—more accurately, ability to do more or less well after an interval of time at an intelligence-test—necessarily reflect cognitive changes in the traditional sense. They may represent changes in motivation, or a different degree of mental organisation and conceptualisation resulting from changes in the range of mental operations (Piaget, 1951), which again would be affected by changes in motivation.

## V.—REFERENCES.

- BROWN, R. R. (1933). The time interval between test and retest in its relation to the constancy of the intelligence quotient. *J. Educ. Psychol.*, 24, 81-96.  
 CLARKE, A. M., and CLARKE, A. D. B. (1953). How constant is the I.Q.? *Lancet*, 265, No. 6791, 877.  
 — ed. (1958). *Mental deficiency: The Changing Outlook*. Methuen.  
 ELWOOD, M. J. (1952). Changes in Stanford Binet I.Q. of retarded 6-year-olds. *J. Consult. Psychol.*, 16, 217-219.

- HONZIK, M. P., MACFARLANE, J. W., and ALLAN, C. (1948). The stability of mental test performance between two and eighteen years. *J. Exp. Educ.*, 17, 309-324.
- MCNEMAR, Q. (1942). *The Revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin.
- MILLER, W. S. (1933). Variations of I.Q.s. obtained from group-tests. *J. Educ. Psychol.*, 24, 468-474.
- PIAGET, J. (1951). *The Psychology of Intelligence*. Routledge and Kegan Paul.
- ROBERTS, J. A. F., and MELLONE, M. (1952). On the adjustment of Terman-Merrill I.Q.s. to secure comparability at different ages. *Brit. J. Psychol. Statist. Sect.*, 5, 65-79.
- RUDOLF, G. DE M. (1949). Retesting of the intelligence-quotient and social age. *J. Ment. Sci.*, 95, 696-702.
- SCARR, E. H. (1953). Changes in the Terman-Merrill I.Q.s. with dull children. *Brit. J. Statist. Psychol.*, 6, 71-76.
- STOTT, D. H., and SYKES, E. G. (1956). *The Bristol Social Adjustment Guides*. University of London Press.
- STOTT, D. H. (1958). *The Social Adjustment of Children*. University of London Press.
- STOTT, D. H. (1959a). Infantile illness and subsequent mental and emotional development. *J. Genet. Psychol.*, 94, 233-251.
- STOTT, D. H. (1959b). Evidence for prenatal impairment of temperament in mentally retarded children. *Vita Humana*, 2, 125-148.
- TERMAN, L. M., and MERRILL, M. A. (1937). *Measuring Intelligence*. Harrap.
- THORNDIKE, R. L. (1933). The effect of interval between test and retest on the constancy of the I.Q. *J. Educ. Psychol.*, 24, 543-549.
- (1940). Constancy of the I.Q. *Psychol. Bull.*, 37, 167-186.
- VERNON, P. E. (1955). The psychology of intelligence and G. *Quart. Bull. Brit. Psychol. Soc. No.* 26.
- VERNON, P. E. (1957). *Secondary School Selection*. Methuen.
- VERNON, P. E. (1958). The relation of intelligence to educational backwardness. *Educ. Rev.*, 11, 7-15.

(Manuscript received, August 5th, 1959.)

# REFLECTIONS ON RESEARCH ON ALLOCATION TO SECONDARY EDUCATION

By R. R. DALE

(*University College of Swansea.*)

**SUMMARY.** Research on allocation at 11+ tends to ignore some aspects of the situation. The concept of failure in the grammar school is too imprecise. When correlations are made more attention should be paid to the distribution of movement between the variables. Prediction research should take account of the difference between the performance of boys and girls in the grammar school, and between girls', boys' and mixed schools. Low correlations may be partly due to bad educational guidance. Streaming in the grammar school leads to a crystallisation of differences and an artificially high correlation. Prediction research might investigate the possible existence of different subject arrays for different social classes.

## I.—THE CONCEPT OF FAILURE IN THE GRAMMAR SCHOOL.

THE writer has been working for some years on problems relating to the selection of university students, selection of teachers, and various aspects of co-education at the secondary level. Here and there this work enables one to look at research on allocation at 11+ from a slightly new angle, and to make various tentative suggestions.

There is a large comparative element in the assessment of failure at the grammar school. It is natural for the teaching staff of a grammar school to class, say, half of its pupils as average, one-quarter as good and the remaining quarter as inferior and 'failures.' Now even if selection were perfect, resulting in a rise in standard, this tendency would continue to work in the same direction. It is reinforced by the technique used by the Examining Boards of endeavouring to maintain a steady standard by passing a relatively fixed percentage of pupils in each subject each year. In this case, also, failures would obviously continue even if selection at 11+ were perfect. When reading research on allocation, one gets an uneasy feeling that this factor is being neglected. Even though such research concerns itself more specifically with the justice of the allocation order of merit, its findings might be couched in such terms as to make allowance for this failure phenomenon. The lack of precision over the 'pass line' at the grammar school is, of course, a handicap to prediction research at both the primary school and university ends. There is also a danger that this imprecision might create—or might already have created—a situation where too little ability is chasing too high an academic standard, at the expense of the pupils' education and maybe health. The adoption of new style objective tests as part of the Ordinary level examinations would reduce this danger, give us a much more dependable idea of the standard from year to year, and at the same time provide a more reliable criterion for allocation and research.

As an extension of this section, one might instance the danger of using correlations for predictive purposes without keeping an eye on the actuarial aspect. Taking the extreme cases it is, of course theoretically possible, in individual schools, for correlation between the entrance examination and performance in school to be perfect and the standard so high that there are no failures (e.g., at Ordinary level); also for correlation to be perfect and the standard so low that all pupils fail. Other not so obvious dangers are, first



that the movement between the two variables may be relatively more concentrated either in the upper or the lower part of the distribution, thus causing a discrepancy between the number of misfits, calculated from the correlation co-efficient taken at its face value, and the number actually occurring. Such concentration of movement is, of course, more likely at the bottom end of the grammar school allocation owing to the greater homogeneity of this border zone group. Second, the extent to which the correlation departs from perfection may be brought about, mathematically speaking, in a number of ways—by a large number of small movements between the variables, by a much smaller number of big movements, or by a mixture of these; the number of misfits will depend upon which of these operations is occurring.

Research might also make further investigations into the influence of major adventitious occurrences on prediction, e.g., prolonged illness and absence, because there is an unconscious tendency for us to blame the allocation examination for failing to diagnose brain tumours, parental divorce, and the deaths of pupils' mothers. It appears to be a tenable argument that if one is assessing the power of the allocation examination as an academic measure one ought not to compare it with perfect correlation, but first deduct from the latter that decrease due to adventitious circumstances.

## II.—PREDICTION BY SEXES.

Many of the allocation studies have made no distinction between the sexes, and have made one predictive estimate for both combined. This would be valid only if it could be shown that they make equal progress in the principal subjects and if their overall performance at the criterion stage were virtually the same. Unfortunately, however, this is not so. Girls in girls' schools fare appreciably better than boys in boys' schools at the Ordinary level of the General Certificate of Education. Nor is this superiority spread out evenly over all subjects; in mathematics and physics there is even a feminine inferiority, though this is not immediately apparent when the percentages of passes are compared. (Sutherland, 1959). The case, therefore, for separate prediction systems for boys and girls should be examined experimentally. To calculate joint correlation co-efficients, often with unequal numbers of the sexes in the experiment, is scarcely a scientific procedure. Even to calculate these co-efficients separately for the sexes and then combine them is not valid at the Ordinary level and is of doubtful validity even at the end of the second year in the grammar school. For instance, though girls often start off at 11+ with a slight advantage, they have increased this appreciably in the percentages of passes in each subject, by the time the Ordinary level is reached, so that in a list of the two sexes combined, the girls would tend to move up and the boys down, producing a lower correlation co-efficient than for either sex separately. For some individual subjects such as mathematics and English literature, the sex differences become so pronounced that research workers need to tread even more warily; here the situation is also bedevilled by differences in educational guidance, as we shall see in the next section.

## III.—PREDICTION AND EDUCATIONAL GUIDANCE.

Nor are things always what they seem. Part of this feminine superiority in the Ordinary level examination is due to the predominance in the grammar school curriculum of subjects which have a large literary element and involve considerable memorisation, as opposed to subjects such as mathematics (most parts) and physics, in which the ability to reason in mathematical terms is so

important. This is a form of educational guidance which favours the girls and is even now being modified by the increasing emphasis on science. Again, part of this apparent superiority is also due to marked differences in educational guidance within girls' schools as compared with boys' schools. A high proportion of girls, particularly the weaker ones, are permitted to drop mathematics, physics and chemistry at the Ordinary level (and often well before) and take biology instead, which better suits their abilities and interests. The corresponding movement in the boys' schools, i.e., the dropping of literary subjects by the weaker boys, has made much less progress. There is also a small age difference in favour of the boys. We are not concerned at the moment with the rightness or wrongness of this guidance, but with an attempt to clarify the position for the research worker.

When correlation coefficients are calculated between the allocation examination and a grammar school criterion, the size of the coefficient is often regarded as indicative of the predictive power of the allocation examination. This is only partly true. Some of the so-called misfits are created by bad educational guidance—or in some schools by the lack of guidance. There is a tendency for pupils to drift into 'Arts' or 'Science' courses because their friends are drifting the same way, or because father wishes it, etc. It has even been known for the bias of the headmaster to cause a similar pressure. Clearly, it is wrong to blame the prediction examination for faults in guidance which cause a student to fail. In order to put things in the right perspective, it must be acknowledged that there is also much careful and good guidance.

#### IV.—PREDICTION AND SEX OF SCHOOL.

In the present educational situation the general level of attainment in a grammar school varies according to the sex of its pupils. If we take the Ordinary level examination results of maintained grammar schools as a criterion, without taking into account the different percentages of each sex taking each subject, the girls' schools are superior in almost all subjects (Tyson 1930, King 1949, Sutherland 1959). This by no means necessarily indicates any innate differences between the sexes in general aptitude for grammar school work, but it does mean that at the moment the success chances of a girl in a girls' school are superior to those of a boy in a boys' school.

So far this section has in part repeated, from a different angle, what was said in Section II. When we come to consider mixed schools, however, we are on new territory. In maintained mixed grammar schools the attainment of the two sexes comes nearer to equality in total performance, the boys being appreciably better than boys in boys' schools, while the girls are slightly below the girls in girls' schools (Tyson 1930, Mathews 1925, Sutherland 1959), though girls' attainment in mathematics is a prominent exception to the latter part of this rule (Cameron 1923). If social class were taken into consideration, however the positions of the two girls' groups in overall attainment might well be reversed though as yet there is no research precisely on this point. It should be noted, also, that girls in mixed schools tend to drop fewer weak subjects and to take the examination at a slightly younger age.

From the above, we see that the prediction at 11+ of the success of a boy at 'the grammar school' may depend on whether this is a boys' or a mixed school. In the case of a girl we cannot say definitely whether she will have a better chance in a girls' or a mixed school, but we can say with some certainty that she is likely to do better in mathematics in a mixed school than in a girls' school. These, of course, are average tendencies and may be reversed in

individual schools by the presence of a good teacher in the girls' school and a poor one in the mixed school.

#### V.—PREDICTION AND THE CRYSTALLISATION OF DIFFERENCES.

This concept is best understood through an actual example. Let us suppose that the top 120 boys in an 11+ examination have been allocated to the one local boys' grammar school. It is a usual custom for a school to place these boys in order of merit in Forms A, B, C, and D. This procedure in itself will create a crystallisation of the difference between the top thirty in Form A and the next thirty in Form B, and so on. In other words, it is a force working against any reduction in the correlation coefficient between the allocation examination and any first, second—or even third year, criterion in the grammar school. Though free movement of order of merit is possible within each form, this movement is impeded (though not prevented) between forms. In addition, since the average level of academic aptitude will be highest in Form A, these pupils will work at the quickest pace and the difference in attainment between them and the pupils in Form B will be increased, and so on down to Form D. Again, this will tend to maintain prediction at a high level if calculated by means of correlation coefficients. The authorities do not say: "We have predicted the nature of your grammar school performance and now we are going to make sure that we are right," but in effect this is what they do without knowing it. It is rather like a specialist in hospital who took great pains to diagnose the diseases of groups of patients and later went round with germ-laden hypodermic syringes to make sure his diagnoses came true.

This argument, therefore, casts some doubt on the validity of estimating the number of grammar school misfits by correlating the 11+ examination with grammar school criteria at the end of the first, second and even third year. As most readers will know, a somewhat similar phenomenon is frequently observed in secondary modern pupils compared with secondary grammar pupils.

#### VI.—PREDICTION AND SOCIAL CLASS.

Though we know something of the relationship between overall prediction and social class, there has been little, if any, investigation of the possible existence of different patterns of success for different social classes in the subject 'array.' For example, the prediction of success for the children of unskilled labourers may be different in English language from that in mathematics. If this is so, the neglect of this factor, or of the sex factor, may have caused some of the slight discrepancies in research findings for which no explanation has been put forward.

#### VII.—THE CRITERION.

Many researchers calculate correlations between the allocation examination and a grammar school criterion which they themselves choose or arrange. They then use these correlation coefficients to calculate the number of misfits in the grammar school. In this way they project onto the allocation examination the whole of the blame for the lack of perfect correlation or in other words perfect prediction. But is it not possible that it is in part the criterion which is faulty—that neither its reliability nor its validity are perfect? If this lack of perfection brings the criterion into a closer relationship with the allocation examination than the true one would have done, it will cause a spurious decrease in the number of misfits; if it moves in the opposite direction it will cause a rise.

## VIII.—REFERENCES.

- CAMERON, A. E. (1923). *A Comparative Study of the Mathematical Abilities of Girls and Boys in Secondary Schools*. Unpublished M.A. Thesis, University of London.
- KING, W. H. (1949). *A Critical Analysis of the Results of a School Certificate Examination*. Unpublished Ph.D. Thesis, University of London.
- MATHEWS, M. (1925). *The Advantages and Disadvantages of Co-education*. Unpublished M.A. Thesis, University of Birmingham.
- SUTHERLAND, M. B. (1959). *Sex Differences in Examination Results*. Unpublished M.S., Queen's College, Belfast.
- TYSON, G. (1930). *Some Apparent Effects of Co-education Suggested by a Statistical Investigation of Examination Results*. Unpublished M.Ed. Thesis, University of Manchester.

(Manuscript received, October 30th, 1959.)



# PERSONALITY ASSESSMENTS AND ACADEMIC PERFORMANCE IN A BOYS' GRAMMAR SCHOOL

By E. ASTINGTON

*(Senior Modern Languages Master, Stockport School.)*

**SUMMARY.** This study has as its origin the observation that over 40 per cent. of grammar school candidates for the G.C.E. of the N.U.J.M.B. pass in fewer than three subjects. The hypothesis is considered that this unsatisfactory position might be improved by taking into account in 11+ selection those personality qualities which make for academic success.

After a review of studies attempting to determine the personality qualities associated with academic success or failure, and after some preliminary experiments, it was decided to use teachers' ratings, class-mates' ratings and a questionnaire as methods of assessment of personality.

In 1953-4 and 1954-5, about 300 boys in five primary schools in a county borough were rated by their class teachers for persistence, independence, dominance, interest, nervousness and emotional stability. Some 700 boys in the maintained boys' grammar school in the same town were assessed in the same years for the same six qualities by the ratings of three teachers, class-mates' ratings and a questionnaire; in addition, introversion-extraversion and sociability were assessed by the questionnaire and sociability also by class mates' ratings.

At all levels, successful boys received significantly higher ratings than unsuccessful boys for persistence, independence and interest; dominance seemed to have no consistent connection with academic achievement; successful boys showed a slight tendency to be considered more nervous but more emotionally stable; successful older boys tended to consider themselves less extraverted and less sociable, than their unsuccessful fellows.

Grammar school teachers' ratings of persistence, independence and interest showed an average correlation of about .60 with performance in internal school examinations and with G.C.E. achievement. They differentiated significantly between groups of G.C.E. candidates, matched on I.Q., but of contrasting G.C.E. performance. Ratings made in the second year of the grammar school predicted G.C.E. performance better than I.Q., but less well than second-year examination performance.

Primary school teachers' ratings of persistence, independence and interest showed an initially low correlation, which increased with time, with the boys' subsequent grammar school achievement. These personality ratings showed higher correlations with G.C.E. performance five years later than did I.Q., but were less good predictors than performance in the 11+ examination. On this evidence, the use of such ratings might be considered as an element in 11+ selection, but they ought not to be used as a substitute for the 11+ examination.

## I.—PERSONALITY ASSESSMENTS AND PERFORMANCE IN SCHOOL EXAMINATIONS.

In 1953-4 and 1954-5, boys of Stockport School, a maintained boys' grammar school, were assessed for different aspects of personality in three different ways: by teachers' ratings, by class-mates' ratings and by a questionnaire, a form of self-rating.

In 1953-4, 678 boys, in twenty-six different forms, ranging from the first year to the lower sixth, were rated by thirty different masters, on a five-point descriptive scale ranging from +2 to -2 for each quality, details of which will be found in the Appendix. Each boy was rated by three masters who were teaching him in that year, so that each boy's total score for each quality ranged from +6 to -6. The average number of forms rated by any master was between

two and three, although some rated as many as four. The ratings were carried out in the second half of the academic year, between January and June, so that every master would have taught the boys he was assessing for not less than five months.

In 1954-5 the experiment was repeated: 731 boys were rated by thirty-two different masters, between February and June, 1955.

Class-mates' ratings were collected at the end of each academic year. The writer visited each form in turn and after handing out blank sheets of paper and warning boys not to write their name on their paper, gave the following instructions:

"Have you ever tried to judge the character of the other boys of your form? Here is how to do it. I am going to read you descriptions of different kinds of boy and then I am going to ask you to write down the names of any boys in your form who are like the boy described. If no boy is like the description, leave a blank. Don't put your own name in any section; it is other boys you are judging. You may name boys who are absent today.

"You can be absolutely honest because you are not going to put your name on the paper. This has nothing to do with school.

"You will find that the descriptions are in sets of four. The first description in each set is of a boy who possesses a great deal of the quality described; the second boy possesses a fair amount; the third has not got the quality at all and the fourth has very little.

"Listen to the description of the first boy. Don't write anything yet."

Details of the character-sketches are given in the Appendix, where it will be observed that they are based fairly closely on the rating-scales used by the teachers. A score of +2, +1, -2 or -1 was awarded to each boy, each time he was mentioned in one of the four categories and the boy's total score was the algebraic sum of these scores, divided by the number of raters, i.e., the number of boys present in the form minus one.

In July, 1954, 678 boys were rated in this way and 731 in June and July, 1955.

The questionnaire was devised by the writer and consisted of sets of ten statements relating to each of the personality traits studied: dominance, extraversion, independence, introversion, nervousness, persistence, sociability, emotional stability, and submission. In the 1955 version, ten questions on interest (i.e., attitude to school) were included. The questions were arranged cyclically and in alphabetical order of traits, so that statements 1, 11, 21, 31, 41, 51, 61, 71, 81, 91, related to dominance and so on. The boy indicated how far he thought each statement applied to himself by marking it +2, +1, 0, -1 or -2. The final score for each trait was the algebraic total for all the ten statements. Specimen statements will be found in the Appendix.

The questionnaire was completed by 664 boys on the morning of the last day of the Spring term in 1954; 716 boys filled in the questionnaire on the corresponding date in 1955.

In each year-group of about 120 boys, a comparison was made between the personality assessments given to the top thirty 'successful' boys and the bottom thirty 'unsuccessful' boys, judged on their performance in terminal examinations in December and July. In every subject which was examined, a grade of A, B, C, D, or E was given, the grades being awarded to the year-group in the proportions of A 10 per cent., B 20 per cent., C 40 per cent., D 20 per cent., E 10 per cent. A grade of A was counted as +2, a grade of B as +1, C as 0, D as -1, E as -2, and each boy's 'achievement index' for the academic year was the algebraic sum of his grades in the December and July examinations.

divided by the number of subjects in which he was examined and converted into a percentage. Boys who were absent for terminal examinations were not considered, nor were boys transferred from other grammar Schools in the course of the year. In the fifth year, the twenty-five or so promoted directly from the fourth form to the lower sixth were counted as successful and the group of thirty was completed by the five or so with the highest achievement index in the fifth year forms. Those who were in a fifth form for a second year were excluded from this study.

The successful group of about 150 boys consisted then of five sub-groups of thirty successful boys, i.e., the top 25 per cent. in each year-group from the first year to the fifth year; similarly, the unsuccessful group comprised the 150 boys falling in the bottom 25 per cent. of their respective year group. In comparisons of the mean score for personality assessments of successful and unsuccessful boys, the 't'-test of significance was applied; the level of significance is indicated in the tables of results.

TABLE 1  
PERSISTENCE.—MEAN SCORES FOR SUCCESSFUL AND UNSUCCESSFUL GROUPS.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings (+6 to -6):						
1953-4	151	+2.14	149	-1.65	+3.79	1%
1954-5	149	+1.86	149	-1.35	+3.21	1%
Class-mates' ratings (+2 to -2):						
1953-4	151	+0.20	149	-0.26	+0.46	1%
1954-5	149	+0.15	149	-0.17	+0.32	5%
Questionnaire (+20 to -20):						
1953-4	143	+7.45	144	+6.78	+0.67	N.S.
1954-5	144	+6.64	145	+5.95	+0.69	N.S.

Significant differences were found for persistence ratings given by both teachers and classmates at all age-levels. With only two exceptions, for class-mates' ratings in 1954-5, these differences were significant at the 1 per cent. level; the two exceptions were significant at the 5 per cent. level.

In the questionnaire scores, a significant difference in favour of the successful sub-group appeared only once, at the 1 per cent. level, for the fifth year sub-group in 1953-4. In no case did an unsuccessful sub-group achieve a significantly higher score than the corresponding successful sub-group.

Examination of the three different assessments made of individual boys revealed that many successful boys gave themselves very low scores for persistence in the questionnaire, whereas they were considered to be highly persistent by three of their teachers and thirty of their class-mates. It may be that such boys set very high standards for themselves.

On the other hand, many unsuccessful boys gave themselves very high scores for persistence, whereas teachers and class-mates gave them low assessments. Such boys may be either deliberately distorting their score or may have very low standards of persistence.

TABLE 2

INDEPENDENCE.—MEAN SCORES FOR SUCCESSFUL AND UNSUCCESSFUL GROUPS.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings :						
1953-4 .....	151	+1.67	149	-0.80	+2.47	1%
1954-5 .....	149	+1.49	149	-0.68	+2.17	1%
Class-mates' ratings :						
1953-4 .....	151	+0.18	149	-0.23	+0.41	1%
1954-5 .....	149	+0.14	149	-0.20	+0.34	5%
Questionnaire :						
1953-4 .....	143	+3.85	144	-3.01	+0.84	N.S.
1954-5 .....	144	+3.95	145	-3.20	+0.75	N.S.

For teachers' ratings, significant differences at the 1 per cent. level appeared for all sub-groups, both in 1953-4 and 1954-5. For class-mates' ratings, in both 1953-4 and 1954-5, the differences for the fifth year sub-groups just failed to reach the 5 per cent. level of significance; all other differences were significant.

In the questionnaire scores, only two sub-groups, the third and fifth years in 1953-4, showed significant differences, at the 5 per cent. level. As in the case of the questionnaire scores for persistence, no unsuccessful sub-group achieved a significantly higher score than the corresponding successful sub-group.

TABLE 3

INTEREST.—MEAN SCORES FOR SUCCESSFUL AND UNSUCCESSFUL GROUPS.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings :						
1953-4 .....	151	+2.84	149	-1.13	+3.97	1%
1954-5 .....	149	+2.27	149	-1.13	+3.40	1%
Class-mates' ratings :						
1953-4 .....	151	+0.42	149	-0.54	+0.96	1%
1954-5 .....	149	+0.30	149	-0.32	+0.62	1%
Questionnaire :						
Used in 1954-5 only .....	144	+7.51	145	+1.99	+5.52	5%

All the differences for all sub-groups in both 1953-4 and 1954-5 were significant at the 1 per cent. level for teachers' ratings and class-mates' ratings.

Two features were of interest in the questionnaire scores. For the first and second year sub-groups, the differences were not significant, but there was a sudden rise in the difference between the means at the third-year level. The highly significant differences, at the 1 per cent. level, in the third, fourth and fifth year sub-groups were brought about by a decline in the scores of the unsuccessful, whereas the scores of the successful remained steady in all sub-



groups. A possible explanation is that consciousness of failure may cause a significant change in attitude to school.

For persistence, independence and interest, then, significant differences between scores of successful and unsuccessful groups appear at all age levels, for at least two different types of assessment, in both years that the experiment was conducted.

The results for other qualities may be summarised as follows:

(a) No consistent relationship was found between assessments of dominance and academic performance. There was no evidence to suggest that unsuccessful groups sought compensation in dominant behaviour for their lack of scholastic success.

(b) There was no evidence that unsuccessful groups were considered, or considered themselves to be more nervous. In fact, the tendency was in the other direction: teachers regarded many successful sub-groups as more nervous, although there was no support for this view from class-mates or from the questionnaire.

(c) Teachers considered some successful sub-groups to be more emotionally stable, but there was practically no support for this view from the other sources of assessment.

(d) There was a tendency for older successful sub-groups to regard themselves as less extraverted than the corresponding unsuccessful sub-groups.

(e) Some successful sub-groups considered themselves to be less sociable than their unsuccessful fellows, but this view was not supported by class-mates' ratings.

## II.—PERSONALITY ASSESSMENTS AND G.C.E. PERFORMANCE OF GROUPS MATCHED FOR I.Q.

In this investigation, the methods of personality assessment previously described were used, but this time the comparison was made between the mean scores of contrasting groups of fifth-formers, each group consisting of just over thirty boys, matched on I.Q., but judged successful or unsuccessful on their performance in the 'O' level G.C.E. examination of the N.U.J.M.B. All were taking G.C.E. for the first time at the end of a normal five-year course. The personality assessments were those already made in the second half of each academic year.

Pairs of boys were matched, to within 5 points of I.Q., from the successful (four or more passes) group and the unsuccessful (three or less passes) group, the successful member of each pair having at least two more passes than the unsuccessful one.

TABLE 4  
DATA RELATING TO GROUPS OF FIFTH-YEAR G.C.E. CANDIDATES, MATCHED ON I.Q.

	N	Average I.Q.	Average No. of passes	Difference
1954				
Successful . . . .	34	120.94 } Moray House	4.91 }	-2.71
Unsuccessful . .	34	120.85 } Advanced 7	2.20 }	
1955				
Successful . . . .	32	119.25 } Moray House	5.53 }	+3.75
Unsuccessful . .	32	119.47 } Advanced 6	1.78 }	

It will be seen that the successful averaged about five passes each, while the unsuccessful averaged only two, a very real difference in achievement.

TABLE 5

PERSISTENCE.—MEAN SCORES OF SUCCESSFUL AND UNSUCCESSFUL G.C.E. GROUPS, MATCHED ON I.Q.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings :						
1954 .....	34	+0.79	34	-1.26	+2.05	1%
1955 .....	32	+1.37	32	-0.97	+2.34	1%
Class-mates' ratings :						
1954 .....	34	+0.15	34	-0.14	+0.29	1%
1955 .....	32	+0.14	32	-0.19	+0.33	1%
Questionnaire :						
1954 .....	29	+5.96	29	+5.76	+0.20	N.S.
1955 .....	31	+6.22	31	+5.68	+0.54	N.S.

Persistence ratings are seen to be independent of I.Q. ; they distinguish significantly between groups of comparable I.Q., but of differing G.C.E. performance. The questionnaire scores show no such tendency.

TABLE 6

INDEPENDENCE.—MEAN SCORES OF SUCCESSFUL AND UNSUCCESSFUL G.C.E. GROUPS, MATCHED ON I.Q.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings :						
1954 .....	34	+0.23	34	-0.56	+0.79	N.S.
1955 .....	32	+0.87	32	-0.22	+1.09	5% (Almost 1%)
Class-mates' ratings :						
1954 .....	34	+0.10	34	-0.15	+0.25	5% (Almost 1%)
1955 .....	32	+0.06	32	-0.12	+0.18	5%
Questionnaire :						
1954 .....	29	+3.03	29	+1.24	+1.79	N.S.
1955 .....	31	+6.22	31	+5.68	+0.54	(10%) N.S.

Apart from the 1953-4 teachers' ratings, these results are as expected. Ratings of independence differentiate reasonably well between the matched groups.

TABLE 7

INTEREST.—MEAN SCORES OF SUCCESSFUL AND UNSUCCESSFUL G.C.E. GROUPS, MATCHED ON I.Q.

	Successful		Unsuccessful		Diff.	Sig.
	N	M	N	M		
Teachers' ratings :						
1954 .....	34	+1.21	34	-0.79	+2.00	1%
1955 .....	32	+1.87	32	-0.87	+2.74	1%
Class-mates' ratings .....						
1954 .....	34	+0.36	34	-0.39	+0.75	1%
1955 .....	32	+0.28	32	-0.32	+0.60	1%
Questionnaire :						
Questions on interest not	included	in 1954.				
1955 .....	31	+5.42	31	+1.29	+4.13	5%

Assessments of interest differentiate clearly and consistently between the matched groups.

As in the main study, there were no significant differences for dominance or emotional stability. Teachers considered the successful groups to be more nervous in both years, the differences being significant at the 1 per cent. level, but there were no significant differences in class-mates' ratings or in questionnaire scores. In 1954, the successful candidates considered themselves to be less extraverted and less sociable than the unsuccessful, the differences being significant at the 1 per cent. level, but this pattern was not repeated in 1955.

Summing up, then, for each of the qualities persistence, independence and interest, significant differences have been demonstrated between the mean scores of at least two different forms of assessment for successful and unsuccessful groups matched on I.Q. No such consistent pattern appeared for any other quality.

The question may now be asked: Has any real connection been demonstrated between persistence, independence and interest on the one hand and academic success and failure on the other?

It is well known that ratings of personality are subject to the 'halo' effect. "I know that X is successful in school, so I'll put him down as very persistent." It might be a typical thought-process. A warning against this 'halo' effect was given in the instructions to teachers and, in the writer's opinion, the team of raters was both highly competent and highly conscientious; they were interested and made many suggestions for improvements in the wording of the rating-scales. In addition, it should be noted that, in a large school, with parallel forms, individual subject-masters may not have a very clear idea of which boys are the most generally successful in their year-group, so that 'halo' effect would be based on performance in one subject only and, in that way, would be limited in its extent. It might have been better to have each boy rated by as many as five masters, but there is a limit to teachers' time and forbearance.

In order to minimise 'halo' effect in class-mates' ratings, the character sketches read out to the boys were deliberately worded to exclude references to school situations as far as possible, although obviously, this was impossible in the case of interest.

In the questionnaire, answers may be a deliberate distortion of the truth or boys may vary considerably in insight into their own reactions and thought-processes and in the amount of conscientiousness with which they answer the questions.

The devices used for personality assessment were, therefore, far from perfect but at least there were three of them and the whole experiment, which related to boys ranging in age from 11+ to 16+ was repeated in two successive years. For persistence, independence and interest, the three different forms of assessment, which were in general agreement, seem to point to a coherent and consistent pattern of relationship with academic success.

A further check is, of course, to consider the accuracy of the predictions of academic performance which can be made from these personality ratings and to compare the predictions given with those derived from other sources. This topic will now be considered.

### III.—PERSONALITY ASSESSMENTS AND THE PREDICTION OF ACADEMIC SUCCESS.

Of the three devices used for assessment of persistence, independence and interest, only teachers' ratings are here considered as predictors, since it would be impossible to make any practical use of either class-mates' ratings or questionnaire scores in primary schools, whereas it has been suggested that primary school teachers' ratings of personality should be used as an element in selection at 11+, and Universities commonly ask for such ratings, in the case of sixth form pupils applying for admission. It seems reasonable, therefore, to see what relationship exists between teachers' ratings of personality and the pupils' academic performance.

As an estimate of the reliability of grammar school teachers' ratings, correlations were calculated between the combined assessments given by teachers for persistence, independence and interest made in 1953-4 with those given in 1954-5 for the four year-groups of about 120 boys each, who were assessed on both occasions. These assessments range from +18 to -18. Product-moment correlations for the two sets of assessments are presented in Table 8:

TABLE 8  
CORRELATIONS BETWEEN TEACHERS' RATINGS, 1953-4 AND 1954-5.

	N	r
1953-4 1st Year ratings and 1954-5 2nd Year ratings .....	118	.506
1953-4 2nd Year ratings and 1954-5 3rd Year ratings .....	120	.657
1953-4 3rd Year ratings and 1954-5 4th Year ratings .....	121	.742
1953-4 4th Year ratings and 1954-5 5th Year ratings .....	117	.720
Average r (Fisher's 'z' technique) .....		.665

There is a reasonable consistency, therefore, in the teachers' ratings. The average correlation coefficient is about that often reported in studies of personality ratings.

Let us now consider the correlations between these ratings and academic achievement for the school year in which the ratings were made. For first, second, third and fourth year boys, the criterion of academic performance was the 'achievement index' already described; for fifth year boys, only



G.C.E. candidates taking the examination for the first time were included and the criterion was the number of G.C.E. passes. As before, product-moment correlations are presented :

TABLE 9

CORRELATIONS BETWEEN TEACHERS' RATINGS OF PERSONALITY AND ACADEMIC ACHIEVEMENT IN THE SAME ACADEMIC YEAR.

1953-4	N	r	1954-5	N	r
1st Year . . . .	119	·599	1st Year . . . . .	123	·576
2nd Year . . . .	123	·491	2nd Year . . . .	121	·540
3rd Year . . . .	124	·753	3rd Year . . . . .	121	·620
4th Year . . . .	121	·685	4th Year . . . . .	121	·640
5th Year . . . .	80	·478	5th Year . . . . .	87	·592
Average r . . . .		·610	Average r . . . .		·590

Grammar school teachers' ratings of personality reach a consistent and fairly high level of forecasting efficiency in predicting performance in internal examinations and, at the fifth year level, in an external examination.

Let us now consider the forecasting efficiency of the 1953-4 ratings in predicting achievement in 1954-5, using the same criteria of academic performance :

TABLE 10

CORRELATIONS BETWEEN TEACHERS' RATINGS OF PERSONALITY, 1953-4, AND ACADEMIC ACHIEVEMENT IN THE FOLLOWING YEAR, 1954-5.

	N	r
1953-4 1st Year teachers' ratings and 1954-5 2nd Year achievement	118	·473
1953-4 2nd Year teachers' ratings and 1954-5 3rd Year achievement	120	·489
1953-4 3rd Year teachers' ratings and 1954-5 4th Year achievement	121	·654
1953-4 4th Year teachers' ratings and 1954-5 5th Year G.C.E. result	87	·324
Average r . . .		·495

There is a decline in the forecasting efficiency of the ratings, a decline particularly marked as far as G.C.E. performance is concerned. The fourth year ratings show a correlation of ·324 with G.C.E. performance, whereas the fifth year ratings give a correlation of ·592. Experience suggests that, under the influence of the imminent examination, many G.C.E. candidates will begin to show increased persistence and a more favourable attitude to school work during the fifth year.

We will now see what happens when we attempt something much more difficult—to predict G.C.E. performance three years ahead. For the second year group of 1954-5, we have teachers' ratings of persistence, independence and interest, we have their 'achievement index' for that year and, in addition, we have the I.Qs. derived from Moray House Advanced 6 Test, taken in June, 1955. We can, therefore, compare achievement index, teachers' ratings of personality and I.Q. as predictors of G.C.E. performance in 1958 :

TABLE 11

CORRELATIONS BETWEEN PREDICTORS, 1955, and G.C.E. PERFORMANCE, 1958.

	N	r
Achievement index, 1954-5 .....	97	·677
Teachers' ratings of personality, 1955.....	97	·385
I.Q. (Moray House Advanced 6), June, 1955.....	94	·214

Teachers' ratings of personality are by no means as good a predictor of G.C.E. performance as achievement in school examinations, although they are better predictors than scores derived from an Intelligence Test.

This is very much the same pattern of results as that in an experiment previously reported by the present writer (*Times Educational Supplement*, 14th Nov., 1958, p. 1649). In that case, for a group of about fifty or sixty boys entering the grammar school, ratings of persistence, independence, interest and other qualities, made on the scales given in the Appendix, by only one primary school teacher before the 11+ examination, were available. Tables showing rank-difference correlations between, on the one hand, performance in the 11+ examination, teachers' ratings of personality and I.Q. as measured at the time of the 11+ examination and, on the other hand, the achievement index for these boys at the end of each academic year in the grammar school, are here repeated and brought up to date :

TABLE 12

1954 INTAKE.—CORRELATIONS BETWEEN PREDICTORS AND GRAMMAR SCHOOL ACHIEVEMENT.

Achievement Index	11+ examination	I.Q.	Personality ratings	N
1st Year G.S., 1955	·658	·449	·121	55
2nd Year G.S., 1956	·574	·460	·208	55
3rd Year G.S., 1957	·607	·387	·262	54
4th Year G.S., 1958	·561	·439	·275	53
5th Year G.C.E. 1959	·467	·159	·303	38

TABLE 13

1955 INTAKE.—CORRELATIONS BETWEEN PREDICTORS AND G.S. ACHIEVEMENT.

Achievement index	11+ examination	I.Q.	Personality ratings	N
1st Year G.S., 1956	·352	·055	·272	55
2nd Year G.S., 1957	·287	—·103	·316	54
3rd Year G.S., 1958	·310	—·106	·218	51
4th Year G.S., 1959	·443	—·223	·254	47
5th Year G.C.E. 1960	·221	—·254	—·071	40

For both intakes, the primary school teachers' ratings of personality are better predictors than I.Q., but less good than 11+ examination positions.

In the bottom line of the table for the 1954 intake at the fourth year level for the 1955 intake, the correlations with the 11+ examination are significant at the 1% level, those with teachers' ratings at the 10% level, while the correlations with I.Q. are not significant, as is to be expected with small, highly selected grammar school groups. In the bottom line of the 1955 table, none of the correlations with G.C.E. performance is significant, but the order of magnitude is the same as for the 1954 intake and as in table 11, which compared predictions of G.C.E. performance made three years previously.

This is an interesting further piece of evidence, suggesting the importance of the role played by persistence, independence and interest in influencing scholastic success. But on such evidence as this, we would not be justified in replacing the 11+ examination by primary school teachers' ratings of their pupils' personalities, although it might be worth while experimenting with such ratings as an element in selection.

For the moment, and until some less crude and more effective method of assessing aspects of personality is devised, the much-decried 11+ examination, imperfect as it is, seems as good a predictor as we are likely to have in this imperfect world.

#### IV.—APPENDIX.

##### (1) TEACHERS' RATING SCALES.

###### (a) *Persistence.*

- +2 Persistent in overcoming difficulties. Extremely steadfast. Never gives up. Decidedly strong-willed and persevering.
- +1 Rather persevering.
- 0 Average for his age.
- 1 Gives up rather easily.
- 2 Decided lack of will-power and persistence. Easily discouraged and gives up at the slightest difficulty.

###### (b) *Independence.*

- +2 Independent of others in thought and action. Never led by others. Can understand a problem better by studying it alone, rather than by discussing it with others.
- +1 Not usually led by others. Will usually stick to his own point of view.
- 0 Neither markedly dependent on others nor independent.
- 1 Sometimes led by others. Will usually defer to another's point of view.
- 2 Markedly dependent on others in making decisions. Waits for others to give a lead. Usually prefers to be with others.

###### (c) *Interest.*

- +2 Shows keen interest.
- +1 Interest above average.
- 0 Average interest.
- 1 Interest below average.
- 2 Not interested. Unconcerned.

##### (2) CHARACTER-SKETCHES USED FOR CLASS-MATES' RATINGS.

###### (a) *Persistence.*

(i) This boy is very good at sticking at a job or a game until it is finished. He never wants to give up until he has finished. He is very strong-willed and persevering. He is persistent in overcoming difficulties.

(ii) Next write down the names of the boys who are next best at sticking at something they have started. They don't usually give up until they have finished. They are fairly strong-willed and persevering.

(iii) Now we are looking for the people who are the poorest at sticking at something they have started. These boys never settle very long at anything. They are easily discouraged and give up at the slightest difficulty.

(iv) Now write the names of the boys who are the next poorest 'stickers.' They give up rather easily and are not very good at sticking at anything they have started.

(b) *Independence.*

(i) This type of boy is very independent. He is never led by others. He tries to work out things for himself. He probably prefers to work on his own. He usually sticks to his own point of view, and is not influenced by other people.

(ii) Now write the names of the boys who are almost as independent as the ones described in the previous section, but not quite.

(iii) This type of boy waits for others to give a lead. He can't make up his mind for himself. He seems to be dependent on others.

(iv) Now write the names of any boys who are almost as dependent as the one described in the previous section, but not quite.

(c) *Interest.*

(i) This type of boy is extremely interested in school work.

(ii) This type of boy is fairly interested in school work.

(iii) This type of boy isn't interested in school work at all. He couldn't care less.

(iv) This type of boy shows very little interest in school work.

(3) STATEMENTS USED IN THE QUESTIONNAIRE.

(a) *Persistence.*

7.—I can work at a hard task for a long time without getting tired of it.

17.—I take a pleasure in overcoming difficulties.

27.—I can keep on at a tiresome task without someone prodding me.

37.—I finish most things which I begin.

47.—I can enjoy a long spell of continuous activity.

57.—I stick at a job even though it seems that I am not getting results.

67.—I return to a task which has stumped me, determined to conquer it.

77.—I usually refuse to admit that I am tired when I am.

87.—I usually refuse to admit defeat.

97.—I will go to any lengths rather than be called a quitter.

(b) *Independence.*

3.—I would rather work out my own plans than carry out the programme of someone else, even of a person whom I respect.

13.—I become stubborn and resistant when others try to force me to do things.

23.—I try to work out things for myself if I am in trouble.

33.—I argue against people who attempt to assert their authority over me.

43.—I prefer to do my own planning alone rather than with others.

53.—I go my own way, regardless of the opinion of others.

63.—I prefer to work on my own.

73.—I would rather go without something than ask a favour.

83.—If necessary, I could get along quite well without friends.

93.—I usually say 'No' when others offer to help me.

(c) *Interest.*

4.—I get on very well at school.

14.—I think that the subjects we study at school are very interesting.

24.—The day passes very quickly when I am at school.

34.—I think that boys who can stay on at school in the sixth form are lucky.

44.—I think that what is taught at school is very important.

54.—I shall stay on at school as long as I can.

64.—I like some school subjects enough to read about them in my spare time.

74.—I am happy when I am learning about school subjects.

84.—I would rather work with my brain than with my hands.

94.—If I could choose, I would rather come to school than go to work.

(Manuscript received, October 19th, 1959.)



# PLACEMENT PROBLEMS AMONG ENGINEERING APPRENTICES IN PART-TIME TECHNICAL COLLEGE COURSES

## PART I.—RANGE OF ABILITY.

BY ETHEL C. VENABLES

(*Nuffield Research Unit, Institute of Education, University of Birmingham.*)

**SUMMARY.** Three groups of part-time day release students taking first year Ordinary National Certificate Classes in engineering were given a verbal and non-verbal intelligence test ( $N > 1000$ ). On the verbal test the mean score was only slightly above that for the population as a whole, but on the non-verbal test 89 per cent. were at or above the fiftieth percentile with 27 per cent. in the top 10 per cent. of the population. Recent attempts to improve the selection of Ordinary National Certificate students appear to be unrelated to intelligence test scores, as the results obtained in two different areas between 1950 and 1957 show no significant differences. Fifty per cent. had attended secondary modern schools as against 73 per cent. of the whole school population in the same age range. Seventy-five per cent. had fathers in skilled and managerial jobs, compared with a national average of 53 per cent. In 1957, only 18 per cent. of those not in full-time education were attending part-time day classes.

The educational and sociological implications of these figures are discussed in relation to manpower problems. An increase in the number of such students must increase the proportion from secondary modern schools and unskilled homes, so that the problems associated with teaching young people of poor verbal ability are unlikely to diminish.

### I.—INTRODUCTION.

THIS paper is the first of three which will summarise the results of a study of part-time day release engineering students in National Certificate and Trade courses in technical colleges.

It deals first with the range of ability to be found in these classes. It has been suggested that it is here that any reserve of ability is to be found.<sup>14</sup> Next it deals with the differences between the verbal and non-verbal abilities of these students.

Until recently, students in technical colleges have more or less escaped the attention of educational psychologists. Originally, the local 'tech' embodied the idea of opportunity for the under-privileged and a second chance for all comers, so that each student was allowed to decide for himself which course or examination he would take. The growth of industry and the increasing need for education has led to a great change within the technical colleges, from a situation in which only highly motivated individuals presented themselves as potential students, to one in which part-time attendance at the local technical college is frequently a requirement of the industrial job. The attitudes of such students range all the way from that of the keen volunteer of the last century to the indifference of the 'practical' man who rejects completely any intellectual approach to the work he has to do. Between these two extremes the young people concerned are likely to be influenced in their attitudes to college studies not only by teachers and classroom companions, but by foremen and managers who may or may not regard academic certificates as a necessary adjunct to practical skill.

In 1952, two years after this research began, 13 per cent. of young people aged 15-17 years, who were not undergoing full-time education, attended technical college classes on one day per week. Engineering industries supplied by far the highest proportion of these students (34 per cent.) and the results reported here refer to students in engineering departments only. In Engineering departments there are two main types of courses—National Certificate and a variety of trade courses leading to Certificates of the City and Guilds of London Institute. In recent years the need for a third stream<sup>12</sup> for those of lower academic ability has been recognised and many colleges now run three types of classes within one department. In an educational system with a fairly rigid selection for grammar school and university, the policy of an 'open door' in further education is too valuable to lose, but the problems of 'placement' in appropriate courses are becoming increasingly difficult.

## II.—METHOD OF ENQUIRY.

Batteries of tests were given to first-year engineering students in National Certificate and Trade courses in 1950 and 1952 in four colleges in the North-West and in 1957 in three colleges in the Midlands. The batteries varied slightly on each occasion but all students were given the N.I.I.P. Group Test 33 and Raven's Matrices (1947 version). (The 1938 version proved too easy and did not discriminate between the two groups of students.)

## III.—RESULTS.

Test scores obtained in 1950 and 1952 in the North-West and in 1957 in the Midlands, showed no significant differences.

Twenty-seven per cent. of students in the first year of an Ordinary National Certificate engineering course were in the top 30 per cent. of the population on the verbal test of ability. On the non-verbal test 77 per cent. had scores at or above the 70th percentile (Table 1).

TABLE 1

PERCENTAGES OF PART-TIME DAY RELEASE ENGINEERING STUDENTS SCORING WITHIN THE SAME RANGE AS THE TOP HALF OF THE POPULATION ON TWO DIFFERENT TYPES OF INTELLIGENCE TEST.

General population	N > 1,000 Part-time engineering students	
	Raven's Matrices 1947	N.I.I.P. G.T.33
Top 10%	37%	7%
Top 20%	60%	16%
Top 30 %	77%	27%
Top 50%	87%	58%

When compared with university student norms, similar differences between the two sets of test scores are evident. Table 2 shows that only 5 per cent. of these engineering students were in the same class on the verbal test as the top 75 per cent. of university students, but according to Table 3, nearly 40 per cent. were comparable when judged by the non-verbal test.

TABLE 2

SCORES ON N.I.I.P. G.T.33 OF FIRST-YEAR PART-TIME ENGINEERING STUDENTS COMPARED WITH UNIVERSITY STUDENTS.

Scores	*University Students	First-year O.N.C.			First-year Trade
		1950	1952	1957	1950/1952 combined
185 and over .....	5	0	0	0	0
184-165 .....	86	1	4	2	1
164-156 .....	(110)	2	0	6	1
155-145 .....	( )	9	14	16	6
144-125 .....	50	45	66	82	29
124-105 .....	11	81	101	191	78
Below 105 .....	0	103	93	216	400
N .....	262	241	278	513	515
Mean .....	156.31	108.18	108.98	108.59	82.97
S.D. ....	16.19	23.18	22.24	21.10	26.43
Within top 75% of Univ. Group (approx.) ....	75%	5.0%	6.5%	5%	1.5%

\*—Graduates in Arts and Science in a University Education Department.<sup>18</sup>

TABLE 3

SCORES ON RAVEN'S MATRICES 1947 OF FIRST-YEAR PART-TIME ENGINEERING STUDENTS COMPARED WITH UNIVERSITY STUDENTS.

Scores	*University Students	First-Year O.N.C.			First-year Trade
		1950	1952	1957	1950/1952 combined.
44-41 .....	2	5	0	4	1
40-37 .....	28	11	7	24	4
36-33 .....	65	37	54	108	18
32-29 .....	53	74	88	154	68
28-25 .....	19	59	68	109	125
24-0 .....	3	55	65	124	295
N .....	170	241	282	523	511
Mean .....	32.9	28.18	28.26	28.24	22.72
S.D. ....	3.9	6.0	6.0	6.1	7.0
Within top 75% of Univ. Group (approx.) ....	75%	37.3%	37.2%	40.7%	11.1%

\* About half this group were third-year medical students and the rest graduates in Arts and Science.

## IV.—DISCUSSION.

During the years covered by this research there has been much discussion about wastage and failure on these part-time courses. Colleges have been attempting in various ways to improve their methods of placing students in the two or three types of courses available; the more academic National Certificate courses on the one hand and the less demanding Trade courses and practical 'Third Stream' courses on the other.

Intelligence test results obtained in the Midlands in 1957 and shown in Tables 2 and 3 are not significantly different from those achieved by students in the North-West in 1950 and 1952, suggesting that the methods of selection in use are unrelated to intelligence test scores and still fail to exclude from the more difficult course those of poor ability who are better suited to trade or Third stream Classes.\*

Nevertheless, the results leave little doubt that among these first-year students there is a large number with good ability. The overlap with university students on the Matrices is considerably higher than on the verbal test. A tendency for Arts students to do better on a verbal test than those from the Science faculty is well established. The test scores in the first column in Tables 2 and 3 are for Arts and Science graduates taken together, which tends to exaggerate the differences: it would be fairer to make the comparison with Science and Technology graduates only, but the full details are not available. Tozer and Larwood<sup>13</sup> found that the mean score on the verbal test for Science graduates only was 148.67 and Bell,<sup>1</sup> testing a group of thirty graduates with Raven's Matrices 1947, found the mean for the Science group to be 33.61 compared with a mean of 32.11 for the whole group. Nevertheless, even when these allowances are made, the difference in the ability of these engineering students as measured by the two types of test is very marked.

Vernon<sup>15</sup> found that a group of training college students whose scores on a non-verbal test (Stephenson's) were similar to those of a group of graduates, made much lower scores on N.I.I.P. G.T.33. He suggested that the difference could be explained by assuming that the two groups were of equal 'g' but that verbal tests "are markedly affected by university training." His recent work with secondary school children suggests that a grammar school education tends to produce a similar improvement in verbal test scores but has less effect on non-verbal ability.<sup>17</sup>

The universities and the grammar schools, are not selecting simply the people with high 'g' but those with high verbal ability as well. Are the grammar school leavers who are considered to be unsuited to an academic course largely those who score very differently on the two types of intelligence test? Or, to express it in a different way, are they those whose verbal ability (or lack of it) is such that the grammar school training has failed to improve (or remedy) it? Table 4 makes it clear that although the ex-grammar school boy is better verbally than the others, there is still a considerable difference between his percentile ranking on the two tests. Only 15 per cent. of them are below the 70th percentile on the Raven's test, but when judged by the Verbal test 40 per cent. of them are in that category. Of the boys and girls entering the grammar schools in 1946, a third of those considered by the schools to be capable of taking an advanced sixth form course left at 16<sup>4</sup>, and more than half of these were considered capable of taking Mathematics and Science at the Advanced level. In the report on "Early Leaving," from which these figures

\* This subject is discussed in detail in the writer's (unpublished) Ph.D. Thesis, Manchester, 1955.



TABLE 4

SCORES ON TWO TESTS ACCORDING TO PREVIOUS SCHOOLING.

Percentiles	Raven's Matrices 1947						N.I.I.P. G.T.33					
	Grammar and Private		Technical		Modern		Grammar and Private		Technical		Modern	
	N	% of group	N	% of group	N	% of group	N	% of group	N	% of group	N	% of group
90th . . . . .	34	46%	37	46%	47	28%	24	32.5%	2	3%	3	2%
70th . . . . .	29	39%	30	38%	70	42%	20	27%	13	16%	21	13%
50th . . . . .	6	8%	7	9%	30	18%	18	24.5%	37	46%	47	28%
Bottom 50	5	7%	6	7%	20	12%	12	16%	28	35%	96	57%
Totals . . . .	74	100%	80	100%	167	100%	74	100%	80	100%	167	100%

are taken, the council also reached the rather surprising conclusion that it seemed likely that "less than half of the present intake into grammar schools could profitably take sixth form work." Even allowing for a large margin of error in grammar school selection, this still leaves a very considerable number of adolescents needing further education if their abilities are not to be wasted. It is also worth bearing in mind that the parents of a small number of ten-year-olds refuse to allow their children to compete for a grammar school place, so that a few secondary modern school pupils may have high verbal ability. Over a quarter of the students tested had scores placing them in the top quartile on the non-verbal test, and in the bottom half of the population on the verbal test.

How far is this difference innate and how far environmental? Parnell<sup>10</sup> has shown that, in physical build, engineers can more often be classified as mesomorphs than arts men, and also that there is some relationship between mesomorphy and non-verbal ability. Non-verbal tests, although heavily saturated with 'g', also have loadings in the factor for space perception. Parnell's results cannot, as yet, be explained, but it is possible that specific abilities (rather than general ability) tend to be linked with physical make-up. Again, physical make-up may play some part in the direction of interests and thus in the choice of a job.

Environmental differences could arise in at least two ways. Lack of verbal stimulation in the early years may lead to a fundamental difference in methods of thinking and response which would be difficult, if not impossible, to reverse. Some sociological work on perception<sup>2</sup> and on social class and educational opportunity<sup>5</sup> might be interpreted in this way.

On the other hand, the differences could arise through a lack of interest and attention. Differences in home conditions may affect early responses to schooling leading to a rejection of academic values and consequent under-achievement. If this is to any extent true, then, when such a youth enrolls in the local technical college, it is not only he who is being given a second chance, but the educational system as well. Such students present the teacher—not only of English, but also of the other subjects—with a challenge, which will have to be met if the problem of scientific and technological man-power is to be solved.

Vernon, including in his list of 'Human Abilities' factor X (which is variously called 'studiousness,' 'interest,' or 'industriousness'), says "Little is known about X and further research would be profitable. In particular, we would like to know how far it depends on (a) home background, (b) the tone of the pupils's school, (c) the stimulatingness of, or good teaching by, his teacher, (d) the pupil's interests, (e) his temperamental characteristics."<sup>16</sup> Vernon has himself produced evidence that the intellect can be 'spurred'<sup>17</sup> though how far his results are due to (a), (b) or (c) it is not possible to say. Lovell<sup>8</sup> in an interesting paper in this field has listed the characteristics of a stimulating school, indicating how (a), (b) and (c) are related and impossible to disentangle. He cites the work of Hearnshaw<sup>7</sup> on 'Concept Formation' and concludes that further research is needed into the problem of whether a "relative inability to categorize is due to poor quality education and limited cultural background or whether it reflects a change in the response of an individual to the environment."

Vernon, summarising his own and Lovell's results<sup>18</sup> concludes that "Certain aspects of intelligence, in particular, flexibility of thinking, are as much or more, affected by the stimulation that different environments and different types of schools provide as are the conventional attainments. Clearly, far more research is needed into the nature of these abilities and into the types of education that best helps to develop them."

The profound effect of the change from school to work is a matter of common observation. That firms vary in their effect on their apprentices has been shown in the recent report of the N.I.I.P. of a validation study of their tests for engineering apprentices,<sup>9</sup> and there can be no doubt that colleges and their staffs frequently have a beneficial influence. Most students attend these courses between the ages of 16 and 21—a time when there can be a great, and often unexpected, development, and when their interests are continually changing. Certainly one of the most rewarding aspects of work in a technical institution is to watch an uncouth, rowdy or shy schoolboy change into a self-respecting and skilled citizen.

Research is in progress in three colleges to discover whether there are any changes in test score during the years of study, and if so, to what extent they are related to examination success.

Viewed nationally the numbers of young people involved in further education classes is very large (230,000 in 1957/58).<sup>9</sup> It may be asked what consequences are likely to follow the increase of day-release students which is envisaged.<sup>11</sup> In the sample tested only half were from secondary modern schools though at that time they catered for 73 per cent. of secondary school children, and only 9 per cent. had fathers in semi and unskilled jobs, whereas 29 per cent. of all workers can be so classified.<sup>3</sup> Hence, assuming the sample to be representative, it seems likely that the increase in students will mean an increased proportion from secondary modern schools and from semi-skilled and unskilled families. Thus it seems likely that the proportion with low verbal ability will also increase.

#### V.—BIBLIOGRAPHY.

- <sup>1</sup> BELL, A. W. (1959). *Unpublished Research*. Birmingham University Institute of Education.
- <sup>2</sup> BERNSTEIN, B. (1958). Some sociological determinants of perception. *Brit. J. Sociol.*, 9, 159-74.
- <sup>3</sup> CENSUS (1951). *1 per cent. Sample Tables*. H.M.S.O.
- <sup>4</sup> CENTRAL ADVISORY COUNCIL FOR EDUCATION (ENGLAND) (1954). *Early Leaving*. H.M.S.O.

- <sup>5</sup> FLOUD, J. (1956). *Social Class and Educational Opportunity*. Heinemann.
- <sup>6</sup> FRISBY, C. B., et al. (1959). *Tests for Engineering Apprentices—A Validation Study*. N.I.I.P. Report 14.
- <sup>7</sup> HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.*, 42, 315-21.
- <sup>8</sup> LOVELL, K. (1955). A study of the problem of intellectual deterioration in adolescents and young adults. *Brit. J. Psychol.*, 46, 199-210.
- <sup>9</sup> MINISTRY OF EDUCATION. *Education in 1958*. H.M.S.O.
- <sup>10</sup> PARNELL, R. W. (1958). Behaviour and Physique. *An Introduction to Practical and Applied Somatometry*. London: Arnold.
- <sup>11</sup> PART, A. A. (1959). Education for industry and commerce. *Journal of the Royal Soc. of Arts*, 108, 16-29.
- <sup>12</sup> RATCLIFFE, R. (1953). The third stream. *B.A.C.I.E. Journal*, 83-85.
- <sup>13</sup> TOZER, A. H. D., and LARWOOD, H. J. C. (1953). An analysis of intelligence test scores of students in a university department of education. *Brit. J. Psychol.*, 44, 347-58.
- <sup>14</sup> EDITORIAL COMMENT (Feb., 1952). *Universities Quarterly*.
- <sup>15</sup> VERNON, P. E. (1939). Educational abilities of training college students. *Brit. J. Educ. Psychol.*, 9, 233-50.
- <sup>16</sup> VERNON, P. E. (1950). *The Structure of Human Abilities*. London: Methuen, p. 38.
- <sup>17</sup> VERNON, P. E. (1955). Spurring the intellect. Letter in *Times Educational Supplement*, 6th May, 1955.
- <sup>18</sup> VERNON, P. E. (1957). Intelligence and intellectual stimulation during adolescence. *Indian Psychol. Bulletin*, 2, Part I.

(Manuscript received, August 22nd, 1959.)

# BEHAVIOUR IN CLASS GROUPS OF CONTRASTING CLIMATE

By D. V. CONNOR

*(Lecturer in Education, University of Otago, Dunedin.)*

**SUMMARY.** In this study classroom groups of varying 'climate' were identified by means of a questionnaire. A comparison of the behaviour of the groups of 'contrasting climate' was made in terms of social interaction, teacher-pupil rapport, social structure, emotional adjustment and attitude to school. An analysis of the results shows: (1) that the class as a social unit seems to have more relevance to the kind of 'climate' found than does the school as a whole, and (2) that the quality of the behaviour in those classes which are identified, on the basis of the original questionnaire, as having 'good climates,' varied widely from those having 'poor climates.'

## I.—INTRODUCTION.

It is now widely agreed that a class must be viewed not only as a group formally set up for learning or instruction, but also as one where a very complex set of inter-personal relationships is operating, which must be taken into account. These dynamic relationships between pupil and teacher and pupil and pupil make for good or poor classroom adjustment and are often referred to as the 'climate' of a classroom. In viewing a classroom group in this way a teacher realizes that, in carrying out his work, he must go beyond his materials and teaching methods and that he has at his disposal forces that are very important for motivating and directing the learning process and for achieving the goals which are originally set.

The important basic elements in any particular classroom climate are the home and school settings and the personalities of the participants as individuals. However, it is the unique conditions which are generated in the teacher-pupil-pupil interactions that constitute the classroom climate proper and it is to these conditions that the present study is directed.

## II.—PREVIOUS RESEARCH.

In earlier studies of classroom climate, it was concluded that the personality and behaviour of the teacher set the pattern for the whole class. For example, Lippitt and White (1943), and Anderson and Brewer (1946) showed behavioural distinctions between groups who were influenced by teachers and group leaders with either authoritarian or democratic approaches. More recently, however, these generalizations have been qualified by demonstrating that there is a fair amount of flexibility or variability in a teachers' characteristic influence depending both upon the individual pupil and the particular learning situation. Another extension of the original studies is the work of Flanders (1951) who relates these earlier findings to a theory of instruction or learning in terms of pupils' goal setting and achieving under different teacher influences. On a wider frame of reference Kemp (1955) has shown that the characteristics of a school, and Campbell (1952) that the cultural level of the home also have an important effect on adjustment and attainment at school.



## III.—PLAN OF THE INVESTIGATION.

The present study is limited to an examination of the relative influences of the school and of the class group itself on classroom adjustment and describes the differing behaviour of groups of contrasting 'climate.' It postulates that wide variations in 'climate' can be objectively demonstrated, that these differences are associated with influences inherent in the class rather than in the school as a whole, and that the quality of the teacher-pupil behaviour in such contrasting classes is consistent with the particular climate revealed.

(1) *Subjects and Materials.* In order to locate groups of contrasting climate, Flanders's\* "Adjustment to School" questionnaire was administered to all the schools containing Standard IV classes in an urban area. This involved 681 children from eighteen classes in eight different primary schools which sampled all socio-economic levels in a metropolitan area. Flanders's questionnaire consists of seventy-eight questions of which the following are a representative sample:

- 1.—Does the class 'play up' when the teacher leaves the room? Yes..... No.....
- 2.—Are you so keen on your work that you always try your best? Yes..... No.....
- 3.—Have you ever been strapped by a teacher? Yes..... No.....
- 4.—Are you afraid when you don't finish your work? Yes..... No.....
- 5.—Does your teacher have lots of fun with you? Yes..... No.....
- 6.—Would you work just as hard if there were no tests? Yes..... No.....
- 7.—Does the teacher make you feel foolish? Yes..... No.....
- 8.—Are you sometimes not sure what to do next in class? Yes..... No.....

Preliminary work with the questionnaire by Flanders showed that it differentiated between well and poorly adjusted groups as judged by Inspectors' ratings, and between classes with authoritarian and permissive teachers as judged by the frequency and quality of verbal output of these teachers.

(2) *Procedure.* Using this questionnaire as a criterion, six 'good' and six 'poor' groups were selected from the eighteen groups and intensively studied as follows:

(a) *Observation Period*, during which a study of 'social interaction' among the children was made. Each child was observed in turn for a fixed period of time and account was taken of the number of 'social contacts' which that child made with other children and with the teacher.

(b) *Teacher-Pupil Rapport*: During a second observation period, Wrightstone's (1951) "Teacher-Pupil Rapport" Scale was applied. This is a rating of the quality of the interaction between the children and the teacher in terms of interest, enjoyment, emotional rapport, physical tension, social relations and reactions to orders.

(c) *Social Structure*: A sociometric test containing questions such as "Whom would you like to sit next to?" and "Whom would you like to take home to tea?" was administered to each of the twelve groups. This permitted an analysis of the 'group cohesiveness' of each class.

(d) *Peer Ratings*, or "Guess Who" test where an assessment is made of the 'reputation' of each member of the class for certain emotional adjustments as viewed by every other member. This consisted of small 'pen pictures' of pupils, and the class was invited to identify one from their own group by means of this thumb-nail sketch. In this way, rankings on such personality traits as

\* As yet unpublished.

perseverance, suggestibility, popularity, etc., were combined in a composite score for 'emotional adjustment.'

(e) *Attitude to School*: Fitt's (1956) Scale was given to assess attitude to school in general rather than toward the particular class. This is the usual Thurstone-type attitude scale of weighted items from which a total score for attitude to school was obtained.

Where possible, the reliability of these techniques, with the exception of the sociometric test, was checked in the present study, or evidence of their reliability with similar age groups was available elsewhere. The reliability coefficients from either of these sources are as follows: Wrightstone's "Teacher-Pupil Rapport" Scale .92; Peer Ratings .83; Social Interaction count .61; and Fitt's "Attitude to School" Scale .70 (split half uncorrected). It was considered that these figures were satisfactory since the study was dealing with groups rather than with individual pupils.

From the above data it was possible to compare the behaviour of the contrasting groups. It was also possible to examine separately both the influence of the class as a group and of the school as a whole on the scores obtained by the children on the "Adjustment to School" questionnaire.

#### IV.—STATEMENT OF RESULTS.

Using the total scores on Flanders's questionnaire as the criterion measure of classroom climate or adjustment, an analysis of variance was carried out to see whether the influence of different schools or classes was contributing to the widely varying scores of individual pupils. The results are shown in Table 1.

TABLE 1  
ANALYSIS OF VARIANCE OF "ADJUSTMENT TO SCHOOL."

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Schools .....	2412.72	7	344.57	.29
Between Classes in the same School .....	11542.16	10	1154.22	(not signif.) 18.97
Within classes .....	40352.08	663	60.86	(signif.)
TOTAL .....	54308.96	680	—	—

In this Table the "within schools" variance has been sub-divided into "between classes in the same school" and "within classes." Comparing the mean squares for the different sources of variance it can be seen that the schools as a whole do not account for a significant amount of the variance whereas that for the classes is very highly significant. This suggests that the reasons for the differences in 'climate' revealed by the questionnaire were not to be found in differences between school settings. Consequently, an earlier plan to examine such matters as school tone and morale, facilities, materials, staff relationships, etc., was not carried out.

It was now clear that it was the class as a unit or the conditions within the class which were largely responsible for the wide differences displayed in classroom climate. An analysis of these differences between the best six and poorest

six classes is shown in Table 2. The ' sign test ' has been applied to the scores of the contrasting groups and the significance of the difference is shown in each case.

By comparing the contrasting classes in Table 2 a general picture of their differing behaviour can be gained :

(a) *Social Interaction.* It can be seen that the classes with ' good ' climates have a significantly higher number of social contacts. This means that the

TABLE 2  
ANALYSIS OF BEHAVIOUR OF THE CONTRASTING GROUPS.

Behaviour Observed or Assessed	Contrasting Groups				Significance *		
	Good Climate		Poor Climate				
	Class	Mean No. of Contacts	Class	Mean No. of Contacts			
1.—Social Interaction . . . . .	a	2.1	u	1.2	P < .01		
	b	1.3	v	1.05			
	c	2.7	w	1.2			
	d	1.9	x	1.07			
	e	2.3	y	1.9			
	f	1.1	z	3.2			
2.—Teacher-Pupil Rapport	Class	Total Score	Class	Total Score	P < .001		
	a	34	u	28			
	b	37	v	30			
	c	36	w	19			
	d	38	x	22			
	e	27	y	16			
3.—Group Cohesiveness . . . . .	f	29	z	14	P < .001		
	Class	Rejects	Stars	Class		Rejects	Stars
	a	15	1	u		18	8
	b	14	5	v		17	5
	c	11	2	w		19	4
	d	9	2	x		15	1
4.—Peers' Ratings . . . . .	e	10	3	y	8	7	
	f	14	4	z	13	4	
	Class	S.D.		Class	S.D.		P < .05
	a	8.92		u	11.96		
	b	10.64		v	10.84		
	c	10.56		w	9.12		
d	8.84		x	9.32			
e	11.04		y	11.48			
5.—Attitude to School . . . . .	f	11.20		z	9.00		
	Class	Mean Score		Class	Mean Score		P < .001
	a	43.2		u	58.5		
	b	46.2		v	61.3		
	c	51.5		w	73.1		
	d	58.8		x	64.0		
e	53.4		y	66.3			
	f	62.4		z	64.9		

\* See Siegel, S. *Non-Parametric Statistics*. McGraw Hill, 1956, pp. 68-75.

children interact socially more or express themselves more freely than in the classes with 'poor' climates during the formal work of the classroom.

(b) *Relations Between Pupils and Teachers.* It is very clear that teacher-pupil rapport is better in the well-adjusted than in the poorly-adjusted groups. That is, according to the ratings, the pupils and teacher interact more; the quality of the interaction is warm and friendly rather than aggressive or competitive; there is more interest and enjoyment; the pupil's role is clear; there is no confusion; children are relaxed rather than tense, confiding and friendly rather than strained and fearful.

(c) *Social Structure of the Groups.* If we define 'stars' as those who received more than 10 per cent. of the total number of possible choices, and 'rejects' as those who were mentioned as being liked least of all, an examination of the results shows that there are significantly more 'stars' and 'rejects' in the poorly-adjusted classes than in the well-adjusted classes. This means that the former have more people at either extreme of 'star' and 'reject' and in this sense their group structure is more 'fragmented' or less 'cohesive' than the well-adjusted classes.

(d) *Emotional Adjustment of the Groups.* Since, by the nature of the rating technique used, a distinction could not be made between the contrasting classes on the basis of mean scores which were zero in all cases, the scatter of the scores was examined. It can be seen that there is no clear-cut difference between the groups in emotional adjustment as measured by this technique. There was also no significant relationship between the scores of individual children on the "Adjustment to School" questionnaire and on this kind of rating on emotional maladjustment. It seems that a child's 'reputation' for emotional adjustment as seen by his classmates is something rather different from his 'school adjustment' as measured by his own assessment of himself on a personal questionnaire.

(e) *Attitude to School.* It can be seen from Table 2 that the mean scores for the 'good' classes are significantly lower\* than those for the 'poor' groups, or that the attitude to school of the former is much better than that of the latter. In other words, good classroom climate seems also to be associated with good attitude to school in general and *vice versa*.

## V.—DISCUSSION AND CONCLUSIONS.

The results indicate that it is possible by means of the questionnaire described above, to identify classes varying from 'very good' to 'very poor' in classroom climate, and that the characteristics of these classes differ widely. In almost every case desirable behaviour is shown in the classes where the 'climate' is good. Children of classes with 'good' climates interact more with one another and with the teacher, have better emotional rapport with the teacher, are socially more cohesive as a group, and have a better attitude to school as a whole. On the other hand, children of classes with 'poor' climates are more rigid and formal, are more distant from the teacher, are more fragmented as a social group and have a poor attitude towards school in general.

There are indications also from this study that the type of classroom climate produced is not so much dependent upon the surroundings or the setting of a school but upon what goes on within the class. This particular result, however may be due to the fact that the original questionnaire unduly reflected the influence of the teacher and did not sample wider sociological influences. There is little doubt, though, that in general the school surroundings and the home backgrounds of the pupils, as well as individual personal attributes of pupils and

\* High scores on this scale indicate poor attitude to school.



teacher, have their influence, but it is the interaction of these in a dynamic classroom situation that has more immediate relevance for the teacher and the learning process.

This study does not offer any evidence on the best way to achieve a good classroom climate. It may well be that a good teacher can create a favourable climate directly by the influence of his personality, and his group should then have all the attributes of a well adjusted class as mentioned above. On the other hand, another teacher may achieve the same end indirectly by concentrating on single aspects of the process and eventually bring about a good classroom climate. Whatever procedure is characteristically adapted, this study has shown that important end products are good rapport between teacher and pupils, a reasonably permissive atmosphere, a 'socially well-balanced' classroom, and a healthy attitude to school in general. The attention of teachers can be drawn to such findings, they may use them as criteria for judging their own classrooms and so be aware of the need or otherwise for evaluating their own position.

ACKNOWLEDGMENTS.—I wish to thank the Director of the New Zealand Council for Educational Research, with whose help the project was carried out, for his assistance in planning the work. I should also like to thank the Otago Education Board for the facilities offered and the headmasters and teachers who so willingly co-operated.

#### VI.—REFERENCES.

- <sup>1</sup> ANDERSON, H. H., and BREWER, H. M. (1946). Studies of teachers' classroom personalities. *Applied Psychol. Monographs*, No. 11, 1-156.
- <sup>2</sup> CAMPBELL, W. J. (1952). The influence of home environment on the educational progress of selective secondary school children. *Brit. J. Educ. Psychol.*, 22, 89-100.
- <sup>3</sup> FITT, A. B. (1956). An experimental study of children's attitude to school in Auckland, N.Z.. *Brit. J. Educ. Psychol.*, 26, 25-30.
- <sup>4</sup> FLANDERS, N. A. (1951). Personal-social anxiety as a factor in experimental learning situations. *J. Educ. Res.*, 45, 100-110.
- <sup>5</sup> KEMP, L. C. D. (1955). Environmental and other characteristics determining attainment in primary schools. *Brit. J. Educ. Psychol.*, 25, 67-77.
- <sup>6</sup> LIPPITT, R., and WHITE, L. K. (1943). The social climate of children's groups, in Barker, R. G., et al. *Child Behavior and Development*. McGraw Hill, pp. 458-508.
- <sup>7</sup> WRIGHTSTONE, J. W. (1951). Measuring the social climate of a classroom. *J. Educ. Res.*, 44, 341-351.

(Manuscript received, Nov. 10th, 1959.)

# SYMPOSIUM: THE DEVELOPMENT OF MORAL VALUES IN CHILDREN

## VIII.—FREUD'S THEORY OF MORAL DEVELOPMENT IN RELATION TO THAT OF PIAGET

By R. S. PETERS

(Departments of Psychology and Philosophy, Birkbeck College, University of London.)

**SUMMARY.** The purpose of this article is to develop some conceptual distinctions in order to explore the relationship between Piaget's and Freud's theories of moral development and to suggest questions to which both theories give rise.

Piaget, it is argued, did little more than show that the distinction between customary and rational morality actually has application. Freud's theory assumed rational morality as an end-point of development but his actual moral theory dealt only with (i) the mechanisms by means of which customary morality gets handed on from generation to generation and with (ii) the reasons why people develop an exaggerated and distorted style of rule-following and never emerge to a more rational type of morality.

Neither Freud nor Piaget, it is argued, have a positive theory about the conditions under which the development towards rational morality tends to take place, except for suggestions in Freudian theory about necessary conditions such as a proper love relationship in the early years. Indeed, there is general ignorance about the results of different methods of passing on rules, methods which need to be carefully distinguished from each other.

It is suggested, finally, that the development in Freudian theory of the theory of the 'autonomous ego' might do much to unify Freud's and Piaget's theories and to suggest fruitful lines of investigation.

### I.—INTRODUCTORY.

In dealing with the vast field of the psychology of morals, Kant's aphorism is particularly apposite: that percepts without concepts are blind and that concepts without percepts are empty. On the one hand there has been a great deal of investigation by psychologists, such as the Hartshorne and May *Character Investigation Inquiry*<sup>1</sup> without adequate conceptual distinctions being made; on the other hand moral philosophers have developed many conceptual schemes which seldom get much concrete filling from empirical facts. The aim of this paper is to explore how psychological theories of moral development might be unified and seen in relation to each other by making certain conceptual distinctions. Such distinctions are necessary for getting clearer about what the theories in fact explain, and for rendering the percepts less blind. I propose to attempt this by setting out Piaget's theory very briefly and by dealing with Freud's theory in relation to it.

### II.—PIAGET'S CONTRIBUTION.

(a) *Exposition.*—Perhaps Piaget's most important contribution was to make explicitly a distinction which people who speak of the psychology of morals are too prone to forget. This is the distinction between what might be called conventional morality and the following of a rational moral code. By 'conventional morality' I mean just doing the done thing, or doing what one is told. If a justification for following a particular rule is asked for, the individual appeals to an authority or to what others do or say is right. Usually, however,

the question of the validity of such a code does not arise, as in a closed, tight-knit, society, where norms tend to be undifferentiated. By a 'rational moral code' I mean one for which the individual sees that there are reasons, which he sees could be otherwise than it is, which he follows more reflectively.

Now Kant, of course, made this sort of distinction when he contrasted the autonomy with the heteronomy of the will; and Piaget, both in his moral theory and in his theory of knowledge, has a Kantian point of view. What he did was to pour into the mould of this conceptual distinction a rich filling taken from observation of children at different ages. He showed that the distinction actually has application.

Piaget studied the attitude of children to the rules of both marbles and morality and found a correlation between them.<sup>2</sup> At the 'transcendental stage' the rule appears as something external and unalterable, often sacred. At the autonomous stage it is seen to be alterable, a convention maintained out of mutual respect which can be altered if the co-operation of others is obtained. Constraint is replaced by reciprocity and co-operation. A lie is no longer just something 'naughty' which adults disapprove of, a prohibition which goes very much against the ego-centric wishes of the child, a command stuck on to a mental structure of a very different order from that of the adult, whose letter has to be adhered to, but whose intention is incomprehensible. It now becomes an action which destroys mutual trust and affection; truth-telling becomes a rule which the child accepts as his own because of the reasons which can be given for it. To summarize the main features of Piaget's contributions:

- (i) Piaget insists that there is something of the sort which Kant described as morality proper, distinct from custom and authoritative regulation.
- (ii) Piaget assumes, so it seems, some process of maturation. There is a gradual transition from one mental structure to another.
- (iii) He assumes that this development in the child's attitude to rules is paralleled by his cognitive development in other spheres—e.g., the grasp of logical relations, and of causal connexions.

(b) *Comments.*—Piaget can be criticized both for what he did do and for what he did not do. Amongst the former type of criticism should be included those of J. F. Morris<sup>3</sup> and D. MacCrae<sup>4</sup> who have maintained that many different sorts of things are included under the concept of 'autonomy' which are not only distinct but which also have not all a high mutual correlation. And, of course, much in general could and has been said about Piaget's methods of investigation.

More interesting, however, are criticisms of what Piaget did not do. For instance, he could have investigated whether the transition through the different stages of morality is just a matter of maturation, or whether it depends on specific social or family or educational traditions. In a lecture which he delivered to UNESCO in 1947<sup>5</sup> he assumes both that value judgments become more equitable with age and that the development depends on general features of the culture. But he makes no attempt to establish this in any detail. E. Lerner<sup>6</sup> has investigated this question and finds a correlation between the sort of development which Piaget outlined and the social status of the parents and the extent to which co-ercive techniques of child-rearing were employed.

Havighurst and Taba<sup>7</sup> also studied the moral attitudes of adolescents in a mid-West town and of children in six American Indian tribes and tried to relate them to methods of training, cultural factors, etc.

Such studies are suggestive but very inconclusive. From them three very general observations could be made about Piaget's work.



(a) Piaget's distinctions provide a useful framework for research. The details, however, of this descriptive apparatus need clarifying and tightening up—e.g., his concept of the autonomous stage.

(b) Given that some such transition sometimes occurs, much more needs to be established about the conditions which favour or retard it. These would include a variety of social factors, but of particular importance would be the techniques for passing on the rules of a society. As far as I know, Piaget says nothing about the ways in which parents and teachers help or hinder children in the transition to the autonomous stage.

(c) Piaget says nothing of the extent to which the relics of the 'transcendental stage' persist in the adult mind, and the conditions which occasion a complete or partial failure of the transition to autonomy in a society where such a development is common, and encouraged.

As Freud, so it seems to me, said a great deal in an indirect way about such matters, it is appropriate to pass to his contribution.

### III.—FREUD'S CONTRIBUTION.

Piaget. I have stressed, explicitly made the distinction between conventional morality and following a rational code. In Freud this distinction is only implicit and the features of a rational code are not explicitly sketched.

Philip Rieff, in his recent book,<sup>8</sup> makes much of what he calls Freud's ethic of honesty and of his uncompromising egoism. He suggests that Freud believed in the generalization of the frankness that is a necessary procedural requirement for psycho-analysis. A man must admit his nature, be quick to detect dishonesty and sham in himself and in others. He must accept his natural needs and have a deep suspicion of 'moral aspirations' such as Freud so often encountered in dealing with middle class women at the turn of the century. Freud's 'education to reality' and 'the primacy of the intelligence,' which he explicitly advocated, go no further than what should be called prudence, rather than morality proper as Piaget understood it. For Freud, on Rieff's view, heralded the advent of psychological man, the trained egoist.

Rieff's account of Freud the moralist, is interesting not only in stressing the cool rationality of Freud's own moral outlook but also in giving a certain interpretation of it. This interpretation is highly disputable. For, although Freud described the principle of impartiality or justice as a cloak for envy,<sup>9</sup> he actually said of himself, "I believe that in a sense of justice and consideration for others, in disliking making others suffer or taking advantage of them I can measure myself with the best people I know."<sup>10</sup> This looks very much like the confession of a rational Utilitarian code such as one could find in Sidgwick—or, indeed, in Piaget. It does not sound like the confession of a man who believed only in prudence or rational egoism.

It is, therefore, difficult to say whether Freud himself subscribed to a rational moral code such as that sketched by Piaget, or only to the cautious prudence attributed to him by Rieff. But, from the point of view of this article, it does not matter. For what I want to begin by stressing is that Freud assumed *some* form of rational code, both in his dealings with others generally and in his therapeutic practice. For the aim of psycho-analysis is to strengthen the ego by making unconscious conflicts conscious and by helping people to make decisions of principle with full cognizance of the irrational sources of their promptings and precepts. It is only his basic assumptions both about the distinction between rational behaviour and being at the mercy of the super-ego, and about the *desirability* of rational behaviour, that make his talk of 'education



to reality' intelligible and his therapeutic practice square with his theory about morals.

As a matter of fact, the distinctions which he made between the ego super-ego, and id, were ways of making the distinction between behaving rationally and behaving in other ways. He equated the ego with reason and sanity; he says it 'tests correspondence with reality', 'secures postponement of motor discharge,' and 'defends itself against the super-ego.' When we ignore the pictorial model which Freud's concepts suggest, we see that their main function is to distinguish between acting rationally when we take account of facts, plan means to ends, and impose rules of prudence on our conduct, in contrast to being at the mercy of the id, when we act impulsively or are driven to act, or being at the mercy of the super-ego when we are obsessed or goaded by the irrational promptings of past authorities.<sup>11</sup>

If we bear in mind this basic conceptual framework we can see, roughly speaking, that Freud's contribution to moral psychology falls under three main headings:

(i) In the theory of the id and the unconscious we find a description of typical occasions when we are deflected from conscious aims, when we cloak up rationalizations as a cloak for following our inclinations, when we act unaccountably 'out of character,' and when we seem to suffer from a general inability to decide between different possible courses of action, to control our impulses, or to carry out intentions however well-meaning. Such investigations throw great light on what might be called the *executive* side of moral action. I do not intend to say anything more about this side of Freud's work, in spite of its intrinsic interest. I propose to concentrate on the extent to which Freud deals with defects on what might be called the legislative or judicial aspects of morals. This implies something amiss with the sorts of rules which we apply to given cases or with the inability to see when rules fit particular cases. Such defects are to be found.

(ii) In Freud's theory of the super-ego.

(iii) In Freud's theory of character-traits.

These I now propose to discuss.

(ii) *The Super-ego.*—The first and obvious point to make about Freud's theory of the super-ego is the extent to which his account of the formation of conscience corroborates Piaget's findings about the child's attitude to rules at the transcendental stage. Freud, it might be argued, went further and showed the mechanisms such as introjection, identification, and reaction-formation, by means of which these externally imposed sacrosanct commands come to be interiorized by the child and the standards adopted of that parent with whom identification takes place. This would explain the perfectly familiar procedure of standards being passed on from generation to generation by contact with the earliest admired figures who exert some kind of discipline and provide a model for the child to emulate. Freud, it might be commented, stressed the 'inner voice' aspect of conscience because he took for granted a way of passing on rules which, as has often been pointed out, was prevalent in a patriarchal society where the father exerted discipline in an authoritarian manner, with a lot of *voice*, and where he was taken as the exemplar of conduct. Freud's theory looked after both the negative notion of taking the voice of prohibition into ourselves and the more positive modelling of our conduct on that of some loved and admired figure.

But a closer look at Freud's theory of the super-ego does not altogether confirm this rather obvious interpretation. Even in the matter of the authorita-

tive voice, Freud showed very little interest on the empirical question of how rules were, in fact, passed on, in whether they were in fact taught with a lot of voice. Indeed, he seemed to take the social environment and educational techniques as more or less constant. He was mainly interested in the mind of the child, in the mechanisms of defence—introjection and reaction-formation—by means of which the child either took the parent into himself, in the form of the ego-ideal, as a safeguard against losing a loved object, or took over the parent's reactions to his sexual wishes as a way of dealing with the danger which they represented. He did not simply use his theory to explain how the traditions of a society are handed on by being taken into the child. For most of his theory was used to explain phenomena which were both different from this and different from each other.<sup>12</sup>

On the one hand, he was interested in the fact that some children develop *more rigorous* standards than those demanded by their parents; on the other hand, he tried to explain the familiar feature of types like the arrogant or humble man who have a picture of themselves, an ego-ideal—what Adler called a 'guiding fiction'—which is quite out of keeping with the traits which they, in fact, exhibit. These phenomena, too, he explained in terms of mechanisms which the child employs to deal with wishes in a social environment which is viewed as being more or less constant. The over-conscientious child is one who has turned his aggression inwards; the ego-ideal is the product of narcissism or outgoing love turned inwards. The obsessional and the melancholic exhibit extreme forms of this type of character. In other words, Freud was here concerned with people who had an exaggerated or distorted style of rule-following. This style of rule-following prevents the development towards a more rational way of following rules such as was sketched by Piaget in his account of the autonomous stage.

In his account of the super-ego, therefore, Freud's theory is an important supplement to that of Piaget. For Freud tries to explain why it is that people, as it were, get stuck at the transcendental stage. Some get stuck as a person might who conforms in a colourless way to the standards that are passed on; others exaggerate or distort the standards in question and are incapacitated for rational rule-following later. Of course, in all of us there is left what Freud called a 'precipitate' of parental prohibitions. Following rules in a rational manner can only be viewed as the end-point of a continuum. My point is that Freud assumed some such end-point and gave special explanations for people who got stuck in strange ways a long way down this continuum.

(iii) *Freud as a characterologist.*—Freud's theory of character-traits, too, supports this sort of interpretation. For here again he developed a genetic theory to explain why it is that another defective style of rule-following which impedes the development of ego strength gets laid down in infancy. The picture of the ego wedged between insistent wishes on the one hand and parental prohibitions on the other is again the conceptual framework for the theory. There are, he held, three main periods at which conflicts are likely to occur, the oral, the anal, and the genital stages. The techniques or mechanisms employed by the child to regulate his dangerous wishes develop into traits of character. Freud suggested<sup>13</sup> that cleanliness and orderliness were reaction-formations against wishes for organ pleasure; interest in money was a sublimation, money being a substitute object; and obstinacy was a continuation of reaction to parental pressures. Similarly sarcasm, scepticism, and food faddism were regarded as oral traits.

I have often been puzzled about what Freud thought he was explaining in this theory. Was he explaining the development of what we call character?

Have the carefulness of the calculating determined business man, which is in accordance with the logic of the situation, and the hoarding of the miser, which is not, a similar explanation, the difference being only one of degree? His theory was that anal character-traits are produced by fixation at this stage of sexuality. Did he assume that those who are not fixated but who pass through this stage in accordance with some assumed norm of development, go forward with a minor imprint of orderliness or obstinacy stamped upon them?

In order to get clear about what Freud's theory explained it is first necessary to make some distinctions. By 'character' we can mean, first of all, the sum total of a person's traits like honesty, considerateness, punctuality, and so on. We speak of giving a servant a 'character' in this sense. Presumably, too, the famous *Character Education Enquiry* of Hartshorne and May was concerned with the investigation of character in this non-committal sense. For little more was investigated than the incidence of traits like honesty, deceit, and so on.

When characterologists, on the other hand, talk of types of character, they usually mean more than just the sum total of traits. They see a certain arrangement of traits which are usually subordinated to one, as in the case of the penurious man. Alternatively, a whole range of traits might be displayed in a distorted or exaggerated manner, as in a pedantic person.

But we can speak of people 'having character' in a third sense, when we speak of a type of consistency which is imposed on other traits by adherence to higher order principles such as those of prudence or justice. Roback, in his classic on the psychology of character<sup>14</sup> spoke of character in this third sense. So also did McDougall, in what he said about the function of the self-regarding sentiment. Freud's concept of the strong ego is very similar.

Now Freud's theory of character-traits is obviously to be seen in the context of characterology. Indeed, Ernest Jones speaks of his *Character and Anal Eroticism* as a contribution to this sort of speculation and notes its literary style.<sup>15</sup> He was surely explaining men who had types of character, men in whom certain traits were so dominant that they provided a unifying pattern. And this, of course, explains the connexion between certain types of character and neurosis. Obsessionals, he claimed, manifest exaggerations of anal traits, just as schizophrenics do of oral traits. Freud, I suggest, was perfectly familiar with the penurious character who crops up over and over again in characterology. His genius consisted in spotting the similarity between this pattern of traits and a type of neurosis and assigning a common cause to both. He was put on to this by the findings of his *Three Essays in the Theory of Sexuality* which just preceded his paper on character-traits.

This type of explanation would account not only for obvious types of character, but also, to a certain extent, for the business man who always tends to be a bit more cautious or more optimistic than the facts warrant. There are standards governing decisions like those of the judge, the business man, and the examiner. The usual assumption is that people learn these standards on some kind of apprenticeship basis; they gradually manage to master the know-how necessary for making rational judgments about such matters. We do not have to tell the Freudian story, surely, about all cases of caution in business, or rigour in marking or severity and leniency of sentence. But when we note a characteristic bias one way or the other, when there is a consistent deviation in one sort of dimension, it then begins to look as if the Freudian type of explanation is relevant.

Of course, it might well be asked what the evidence is for Freud's speculations anyway. That is another matter, with which I am not here concerned. For it surely presupposes the prior question as to what the theory was meant to



explain. I have tried to show that it is to be seen as explaining only types of character, not people who *have* character. Freud's concept of the strong ego would cover what is usually meant by 'having character' together with hints thrown out by people like Abraham about the 'genital character' which describes roughly what I have previously referred to as the end-product of Freud's assumed continuum.<sup>16</sup> Such a person would be one who has not become fixated at any of the previous stages and who has passed through in a relatively smooth way to a more rational way of following rules in accordance with higher order principles like those of prudence, impartiality, and consideration for others which I mentioned previously as characterising Freud's implicit assumptions about a rational moral code.

#### IV.—FREUD'S OMISSIONS.

Part of Freud's moral psychology, then, can be considered a supplement to that of Piaget in that, both in his theory of the super-ego and of character traits, he gives special explanations of why people do *not* attain the rational level of rule-following. I have argued that in his implicit assumptions about morality and in his concept of the ego he assumes a norm of development towards what might be called 'having character' in an autonomous way—regulating behaviour intelligently, consistently, with integrity and without self-deception in accordance with principles of prudence, impartiality, respect for others, and so on. In his special theories he explains how it is that people fail to mature in this way because they get fixated at some earlier period and develop a type of character or because they come to adopt rules in an exaggerated or distorted way at the period of super-ego formation—Piaget's transcendental stage.

But as far as I know there is no positive theory in Freud of the conditions under which this desirable development towards rationality tends to take place. There are, of course, suggestions about very general necessary conditions—e.g., a proper love relationship in early life with a mother-figure, as stressed out later by Bowlby.<sup>17</sup> To be deprived of this is likely to lead to distractibility, unreliability, and lack of self-inhibition which are almost definitions of having *no* character, both in the sense of colourless conformity with standards and of autonomous self-direction. But there is no positive theory about *techniques* of child-rearing and the passing on of rules which tend to favour rational development, which Freud treats almost as a maturation norm.

Are there, for instance, certain techniques of child-rearing which tend to create 'fixation'? On Freudian theory 'fixation' is usually regarded as being the product of too much or too little organ pleasure at the early stages. And there are, of course, many cross-cultural studies of weaning and toilet training.<sup>18</sup> But what is the evidence about techniques for the passing on of rules at the later stages covered by Freud's theory of the super-ego and Piaget's theory of the transcendental stage? Plenty is known about methods of child-rearing in the tribes studied by anthropologists; but there is very little established knowledge about what goes on in the various strata of European society. A start has been made along these lines in America<sup>19</sup> and A. N. Oppenheim is at the moment engaged on an ambitious project to remedy just this defect in our empirical knowledge about our own society.

It is pretty pointless to try to rectify our ignorance on such matters by appealing to conditioning theory and to far-fetched analogies between the behaviour of animals and men. For there are many quite different ways of passing on rules to children of which conditioning is only one. So conceptual confusion is only increased if these different ways of passing on rules are not



distinguished as a preliminary to investigating what the result is of using different techniques at different ages. The type of instrumental conditioning used on animals, for instance, must be distinguished from that form of training in which rewards and punishments are promised and threatened and not just administered. This, in its turn, must be distinguished from the use of authority which presupposes the use of the voice, but, in giving commands rather than in promising and warning. Then there is imitation, suggestion, and learning 'by experience' as well as rational instruction. Given such initial distinctions questions like this could be asked: to what extent do various techniques for passing on rules, when used at early ages where rational explanation and instruction are out of the question, incapacitate the child for rational instruction at later ages? If Piaget is right in saying that there is a correlation between a child's attitude to the rules of marbles and to moral rules as he develops, can any generalizations be made about ways of learning rules which, in general, help or hinder the development towards the more rational stage?<sup>20</sup> Luria, when he recently visited England, reported experiments which he was doing on the use of different techniques for passing on manipulative skills at different ages. He has been studying the use of the authoritative voice, the use of instruction with an explanation, and so on, in these very limited contexts. Could anything be done to supplement Freud's and Piaget's work along these lines? Do some techniques—e.g., conditioning—lead to break-downs, blocks and compulsive habits at later stages? Do they encourage types of character rather than help to develop character? And what sort of training leads to the colourless conformity of the man who never really emerges from Piaget's transcendental stage?

#### V.—DEVELOPMENT OF FREUD'S THEORY.

These omissions in Freud's theory are connected with a more general weakness—the lack of any detailed theory of the development of the ego, of rationality in general of which morality is a particular case. In the main those who have followed him, like Melanie Klein<sup>21</sup>, have revised some of the details of his theory about the development of the super-ego—for instance, the age at which it is formed. Others like Horney<sup>22</sup> and Fromm<sup>23</sup> have criticized his theory on the grounds that it is too biologicistic and takes too little account of social differences, or that it treats infantile sexuality and the mechanisms adopted to deal with it as determinants of, rather than occasions for, the development of more general orientations.

There are, however, two other lines of development which are particularly relevant to the thesis of this article. There is, firstly, Money-Kyrle,<sup>24</sup> who claims that Freud's treatment of moral psychology is one-sided because he dealt only with the authoritarian conscience. There is also the humanistic conscience which is traceable back to the 'guilt' experienced in hurting and hating the mother, the first object of a child's love. If the child learns this first lesson, that the good and the bad are often different aspects of the same object, and if he fears lest what is good may be injured by his own efforts or omissions, he has the beginning of a conscience which is more rational and realistic. Whatever is to be made of Money-Kyrle's genetic speculations, which owe much to Melanie Klein, his theory is interesting in that he sees the connexion between rationality and what he calls the 'humanistic' conscience and in that he tries to connect different types of 'conscience'—e.g., that of the obsessional, the hypomanic, and the hypo-paranoid, with different types of falling off from rationality.

There is, however, a second and even more relevant line of Freudian thought represented by Erikson,<sup>25</sup> Hartmann,<sup>26</sup> and Rapaport,<sup>27</sup> who have developed the theory of the autonomous ego which was rather embryonic in

Freud's writings. Erikson has tried, in very general terms, to map the development of the autonomous ego and to make generalisations about conditions which favour or hinder it. Rapaport has tried to sketch a conceptual model which does justice to the autonomous ego, to show the conditions necessary for the development of autonomy, and to connect his theory with other findings, such as those about stimulus deprivation, hypnosis, 'brain-washing,' and Piaget's views on the origins of intelligence in children. This line of development is still exploratory. But a conceptual scheme based on it might help to unify the theories of Piaget and Freud in general as well as in the moral field.<sup>28</sup> It might, for instance, help to clarify Piaget's conception of the autonomous stage. It might also help to connect their investigations with others in the field, and cast some light on what particular investigations in it actually explain; for in this ill-explored field of moral development what is needed more than in almost any other field in psychology is a combination of concrete investigations with conceptual clarity. And this is the point from which my paper started.

#### VI.—REFERENCES.

- <sup>1</sup> HARTSHORNE, H., and MAY, M. A. (1930). *Studies in the Nature of Character* (3 vols.) New York: Macmillan.
- <sup>2</sup> PIAGET, J. (1932). *The Moral Judgment of the Child*. London: Kegan Paul.
- <sup>3</sup> MORRIS, J. F. (1958). The development of adolescent value-judgments. *Brit. J. Educ. Psychol.*, 28.
- <sup>4</sup> MACRAE, D. (1950). *The Development of Moral Judgment in Children*. Ph.D. Thesis, Harvard University.
- <sup>5</sup> PIAGET, J. (1947). *The Moral Development of the Adolescent in Two Types of Society, 'Primitive and Modern.'* UNESCO (quoted by J. F. Morris, *op. cit.*).
- <sup>6</sup> LERNER, E. (1937). *Constraint Areas and Moral Judgment of Children*. Menasha, Wisconsin.
- <sup>7</sup> HAVIGHURST, R., and TABA, H. (1949). *Adolescent Character and Personality*. New York: Wiley.
- <sup>8</sup> RIEFF, P. (1959). *Freud, the Mind of the Moralist* (Chaps. IX and X). New York: Viking Press.
- <sup>9</sup> FREUD, S. (1949). *Group Psychology and the Analysis of the Ego*. London: Hogarth Press, pp. 86-8.
- <sup>10</sup> JONES, E. (1955). *Sigmund Freud, Life and Work*. London: Hogarth Press, Vol. II, p. 464.
- <sup>11</sup> For further development, see PETERS, R. S. (1958). *The Concept of Motivation*. London: Kegan Paul, Chap. III.
- <sup>12</sup> For classic exposition of the theory of the super-ego and ego-ideal, see FLUGEL, J. C. (1945). *Man, Morals and Society*. London: Duckworth, Chaps. II-VI.
- <sup>13</sup> FREUD, S. (1942). *Character and Anal Eroticism*, in *Collected Papers Vol. II*. London: Hogarth Press.
- <sup>14</sup> ROBACK, A. A. (1928). *The Psychology of Character*. London: Kegan Paul, Chap. XXV.
- <sup>15</sup> JONES, E., *op. cit.*, p. 331.
- <sup>16</sup> ABRAHAM, K. (1949). *Selected Papers on Psycho-Analysis*. London: Hogarth Press, Chaps. XXII-XXV.
- <sup>17</sup> BOWLBY, J. (1953). *Child Care and the Growth of Love*. London: Penguin Books.
- <sup>18</sup> See, for instance, WHITING, J. W. M. and CHILD, I. L. (1953). *Child Training and Personality*. New Haven: Yale U. P. and A. Kardiner's various works.
- <sup>19</sup> SEARS, R., MACOBY, E., and LEVIN, H. (1957). *Patterns of Child-Rearing*. New York: Row, Peterson.
- <sup>20</sup> Similar issues are raised and discussed in: HARDING, D. W. (1953). *Social Psychology and Individual Values*. London: Hutchinson.
- <sup>21</sup> KLEIN, M. (1948). *Early Development of Conscience in the Child*, in, *Contributions to Psychoanalysis*. London: Hogarth Press.
- <sup>22</sup> HORNEY, K. (1949). *New Ways in Psycho-analysis*. London: Kegan Paul.
- <sup>23</sup> FROMM, E. (1948). *Man for Himself*. London: Kegan Paul.
- <sup>24</sup> MONEY-KYRLE, R. (1951). *Psycho-analysis and Politics*. London: Duckworth.
- <sup>25</sup> ERIKSON, E. (1950). *Childhood and Society*. New York: Norton.
- <sup>26</sup> HARTMANN, H. (1958). *Ego-Psychology and the Problem of Adaptation*. London: Imago.
- <sup>27</sup> RAPAPORT, D. (1958). The theory of ego-autonomy: a generalisation. *Bull. Menninger Clinic*, 22, 13-35.
- <sup>28</sup> RAPAPORT, D. (1951). The conceptual model of psycho-analysis. *J. Personality*, 20. See also Rapaport's, *The Psychoanalytic Theory of Motivation*, in the 1961 Nebraska Symposium on Motivation.

# METHODS FOR ASSESSING THE VALUE OF EXHIBITIONS

By D. SHEPPARD

*(The Social Survey, Central Office of Information.)*

SUMMARY. Some who design or use exhibitions to convey information are not aware of the research done on their effectiveness. A discussion is given of the various methods which can be used for such studies and a hypothetical example is given to illustrate their use.

## I.—INTRODUCTION.

WHILE the enthusiasm for visual methods of teaching has by now somewhat died down, this is probably due more to the dictates of fashion than to the results of experimental tests on the value of such methods. Nevertheless, quite an amount of data has been obtained on this topic, not only concerned with exhibitions, but also with visual methods for teaching skills (Vernon 1946, Laner and Vernon 1957) and with the effectiveness of diagrams and graphs of the kind used to illustrate written factual reports (Vernon 1950, 1951, 1952).

Exhibitions will be the main concern here, using this term to include museums, government exhibitions, displays at agricultural shows and exhibitions on special topics prepared by commercial organizations, or by learned societies. As stated, quite an amount of research has been done on exhibitions, but because of the diversity of exhibitions studied, it is difficult to form any generalized conclusions from them (Sheppard 1959). It follows that a discussion of the methods used should be valuable, since the exhibit designer will have to conduct his own studies if he wishes to assess the value of his efforts. It is relevant to note that such handbooks as are available, on ways of preparing exhibits, seem to be founded less on the results of research than on the opinions of 'experts'—opinions which are sometimes demonstrably incorrect. A handbook which does seem to have taken some research on these topics into account is one prepared by the Food and Agriculture Technical Information Service of O.E.E.C. (1958).

Before discussing methods for evaluation, differences between exhibitions should be considered. A main difference is in their purpose: a museum, or a government-sponsored exhibition, is usually designed to educate, while a commercial exhibition is usually intended to induce some particular action such as the purchasing of goods displayed. Some exhibits have a function lying between these two extremes, as for example displays prepared by the Ministry of Agriculture, Fisheries and Food, which are intended to advise farmers on techniques, although not to sell them specific branded products. Few exhibitions are intended only for amusement, except those on piers at the seaside.

A further main difference is in the type of visitor. Exhibitions such as the Ideal Home Exhibition are intended for the general public, while exhibitions such as that held by the Physical Society are intended for specialists. Museums and agricultural shows serve both the specialist and the general public, although schemes such as charging higher entrance fees for certain days may be used to segregate these two groups.

One more difference is in the duration of such exhibitions. A display in a museum or art gallery may remain unchanged for years, while an agricultural show such as 'The Royal' lasts only for four days in each year.



Turning to methods for evaluating exhibitions, it will be best to list these and to give some discussion of the merits, or limitations, of each. An example will then be given of how a particular exhibition might be studied.

## II.—THE METHODS.

Something can be learnt about the value of an exhibition by counting the number of tickets sold, the number of catalogues sold, or the number of leaflets taken away. General indications of such a nature can be useful when the exhibition concerned is a regular event, to see if it is becoming less, or more, popular; to give ideas as to the best site to use, or to compare different ways of advertising it. To make more detailed deductions from such evidence is unwise, an example being a report by Knee (1953), in which he states that the value of Ministry of Agriculture stands at shows can be deduced from the fact that 60,000 technical bulletins were sold to farmers at such shows. In practice, there was no way of knowing if all these bulletins were sold to farmers, and it would be expected that many who bought bulletins would have purchased more than one. Furthermore, while such a figure gives some guide as to the value of the bookstalls, it provides no evidence at all as to the value of the many other exhibits at these stands. Counting of a somewhat more elaborate nature can be useful. Thus counts can be made, at intervals, of the numbers of visitors looking at particular exhibits within an exhibition. Sheppard, Smith and Carrinci (1960) used this method at agricultural shows but found it to be unsatisfactory since it confounds the number of visitors looking at an exhibit with the time spent looking at it. For example, counts at a bookstall made it appear that this was more popular than most other exhibits, but this was because visitors spent a long time there. The number of visitors who stopped to look at the bookstall was, on the average, no greater than for other exhibits.

A further method consists of observing individual visitors as they go round the exhibition. Thus the observer may note which exhibits a visitor looks at, the order in which he looks at them and the time he spends at each. Such observations have been made in a number of investigations (Marston 1925, Robinson 1928, Willcock 1949a, b, Orlansky 1951, Brooks and Vernon 1956, Sheppard, Smith and Carrinci 1960), one suspects in some cases because the investigators were apprehensive about visitors' reactions to being questioned. The method is useful and objective in the sense that there should be practically no subjective element in the data obtained, the observers' role here being little more than that of a ciné-camera. The visitor can usually be observed unobtrusively, so that it can be assumed he is behaving naturally. Check questions on this at agricultural shows indicated that only 3 per cent. of the visitors who had been observed had noticed that this was being done. The method has three main disadvantages: a long time is required for each visitor, so that it is not always possible to collect data for any considerable number. No evidence is provided as to the visitor's occupation, which can be important where the investigator is primarily interested in a particular section of the visiting public: examples are the agricultural show studies, where the main interest was in farmers. Finally, the method gives no indication of the effect of an exhibit, time than is needed to read or see all that is displayed there. 'Standard' times can readily be obtained for this purpose, by timing people who have been instructed to see exhibits properly (Willcock 1949a, b, Sheppard, Smith and Carrinci 1960).



A variation on this technique is to observe visitor behaviour at a few selected exhibits of particular interest to the designer. Less time is required per visitor if this is done, though it is not then possible to make a generalized statement about all the exhibits, nor to compare the attractiveness of the exhibits studied with all the others. Otherwise, the same advantages and disadvantages apply.

A fixed camera can sometimes be used instead of an observer, when the exhibition is concentrated in a fairly small space and is reasonably well lit (Nielsen 1946). It then becomes possible to make detailed studies, unobtrusively, of several visitors at a time. However, at crowded periods, it may prove impossible to identify and trace the course taken by selected individuals (Nedzel 1952). This could well bias the results considerably, since it means that no such data can be obtained at times when, ideally, larger numbers of visitors should be included in the sample; these would also be times at which visitors would be expected to behave differently, just because the exhibition is crowded.

This leaves interviewing to be considered. There is little need to be apprehensive about doing this, for visitors approached at exhibitions rarely refuse to be interviewed. Refusal rates vary from 2 per cent. to 7 per cent. (Sheppard, Smith and Carrinci 1960, Orlansky 1951). Interviews need not be confined to the exhibition itself, and studies of visitors before they go to an exhibition, or after they have left, can provide very valuable data.

Only by interviewing can one assess the effect of the exhibition on the visitor's attitudes, interests, knowledge or activities. On the other hand, the assumption is implied that the visitor usually knows what he has seen and what effects the exhibition has had on him, and that he is prepared to be quite frank and honest about such matters when he does know. The possibility always exists that he will seek to justify his actions in coming, or that his responses will be influenced by his judgment of the status of his interviewer.

About the best method so far devised is to observe visitors as they look at exhibits and to follow this up by interviewing the same visitors as they leave the exhibition (Willcock 1949a, Orlansky 1951, Sheppard, Smith and Carrinci 1960). This method can provide information about all exhibits seen by a visitor; a visitor's statements about what he saw can be compared with the results of the observations, and all such data can be related to his occupation or specialized interests. Some of the disadvantages which apply individually to observations and interviewing still apply when these methods are combined, while this technique has the added limitation that it takes an even longer time for each visitor than straight observations.

There are two further methods which, because of the sampling inadequacies inherent in them, cannot be used for a serious investigation. The first consists of noting remarks made by visitors who are looking at the exhibits and the second consists of asking questions of visitors who, on their own initiative, approach a demonstrator for help or advice. In neither case are those concerned likely to be typical visitors. However, such methods may be useful in a pilot investigation, when searching for clues to be followed up in a later and more extensive study.

Sampling is a key problem in all such investigations, and a brief discussion of this is necessary. Random sampling is rarely feasible, but a method which can be used is that of sampling visitors at fixed intervals. Thus, counts may be made of visitors during every alternate half-hour period, or every third visitor to leave the exhibit may be interviewed. With observations or interviews, however, it is difficult to use this fixed interval sampling method when the

attendance rates vary considerably with time of day. The most usual alternative procedure is for each observer to select, say, the third person to enter the exhibition after the last person who was observed has left. Interviewers may follow a similar procedure (Willcock 1951). If counts of the numbers of visitors who see the exhibit are also made, the data obtained can be weighted so as to take account of a low sampling fraction, when necessary.

While such a sampling method does not provide a truly random sample, the bias introduced is probably not great and, at least, the method does not allow the observer's, or interviewer's, prejudices to influence the choice of visitor.

One special problem arises when the survey is to be concerned only with a particular group of visitors. If these can be identified from their appearance, a procedure similar to that described will give a reasonable sample, using the selection procedure among those visitors with whom the study is concerned and ignoring others. Usually, appearance is not a sufficient guide and yet an unbiased sample has to be obtained. Here, provided that the original method for sampling all visitors was a fair one, the irrelevant data can be thrown away and the data for the sub-sample can be used with confidence that this group of visitors was selected in an unbiased manner. Although such a sampling method for selecting a sub-group may be necessary, it can be frustrating when it leads to extended observations being made of a visitor who is not a member of the all-important sub-group.

A more serious worry is that any generalization will have to be made from a small total number of visitors, which means that averages are then subject to a large sampling error. Small numbers seem inevitable in such studies, for to use too many investigators means that visitors will notice them and perhaps look round the exhibition differently in consequence.

### III.—AN EXAMPLE.

Suppose that a commercial firm wishes to find out if one of their stands held at an agricultural show, is 'effective.' Suppose further that the firm concerned is generous with funds and, despite pressure, those responsible refuse to define what is meant by 'effective.' They have, however, decided on what their exhibit is to be about and on how to display what they are showing, so that there are no possibilities of experimentation with display methods. How would an investigation on this topic be planned?

The first step is to define 'effective' and a possible definition would be the total sales of the products concerned resulting directly from the exhibition. As another measure, the number of persons to whom a sale is made could be considered, either as an absolute number, or as a percentage of all those who came to the exhibit, or of all those who knew that such an exhibit was to be held. Because such a criterion may be difficult to assess, evidence will be obtained on as many aspects of 'client' behaviour as possible: thus, attempts will be made to discover what proportion of the potential visitors visit the show; what proportion come to the firm's stand; how many of these see each individual exhibit on the stand; how many understand, learn or believe what is shown there; how many can recall what was shown, some weeks later, and how many make a purchase as a result.

Having persuaded the firm's representatives to define their 'audience' as farmers and to limit the area they care about to the county in which the show is held, it is decided to do a sample survey at a random sample of farms, after the show is over. This will tell fairly precisely what number of farmers in

this area went to it. However, if few went to the show, such a survey would provide little other useful evidence except perhaps some reasons given by potential visitors for not going.

The investigations have to be planned before the show takes place, so the possibility must be considered that this sample survey will produce few farmers who went to the show, fewer still who attended the firm's stand, and almost no one who purchased the firm's products. Consequently, it is decided to do some studies, at the show, of visitors to the stand, using sufficient investigators to allow a small number (say five) to be on duty throughout the day. One is detailed to make counts of visitors entering the stand, two to interview visitors leaving the stand and two more to observe visitors going round the stand and afterwards interview them.

The counts will provide data on the number of visitors, and the observations followed by interviews will show with certainty what exhibits were seen by farmers. The proportion who learn anything from the exhibits can be assessed by asking the visitors to recall what they saw. More data on learning can be obtained by examining the larger number of interview schedules completed by those who are interviewing only, for these will have asked visitors which exhibits they saw, and what they learnt from those they claimed to see. Somehow, an attempt will be made to allow for the fact that some visitors will not really have learnt anything, but will have been able to guess what was on the exhibit from their general knowledge of the topic. It will be important to know this if any appreciable number of visitors are able to say what was on the exhibits though, of course, it will matter little if only 1 per cent. or so appear to have learnt anything.

Since addresses will have been obtained from the farmers interviewed, it is an easy matter to call on them, or perhaps write to them, later, to see if they have since purchased a sample of the firm's products. If they have done so, they can be asked if this was a direct result of seeing this product exhibited, though that is not to say that such a question can always be easily answered. The farmer concerned may not know precisely what influenced him, or his purchase may have been a result of cumulative influences.

The counts of visitors will provide some basis for re-weighting the data, both to allow for the variability of the sampling fraction and to enable percentages to be expanded into absolute numbers: the firm's main problems can thereby be answered satisfactorily. Let us suppose, however, that the results turn out to be disappointing. The firm, therefore, wants to know why only 3 per cent. of those who came to their stand purchased one of its products. The data obtained show that their exhibit failed because many farmers did not learn what was shown—but was this because these exhibits were incomprehensible? Likewise, those who learnt what was shown did not put into practice what they learnt—was this because they did not believe in the firm's recommendation, or because they forgot what was shown?

It is now too late to measure comprehension at this show, but perhaps at a future show an investigator could accompany certain visitors as they go round the exhibits and ask what they understand from them, or a 'classroom experiment' could be done, showing such exhibits to a sample of farmers, and matching their characteristics with those of farmers who attend shows, by methods such as the B.B.C. have used (Belson 1956). This would show if a particular exhibit could be understood by most farmers. If so, the amount of time needed for full understanding could be compared with the amount of time spent in looking at this exhibit under typical show conditions. An alternative method, which has been tried, consists of asking visitors to shows if they think that the



exhibits could have been made more easy to understand. Visitors tend to regard this as a potential slight on their intelligence, and not as a reflection on the exhibit designer's skill, so that data thus obtained are of little value (Sheppard, Smith and Carrinci 1960).

Belief is more troublesome to investigate. Satisfactory data on this can scarcely be obtained at the stand, but an (apparently) independent survey at the farms concerned might throw light on this. The situation may be complex, for the firm itself may be judged to be untrustworthy, or it may be regarded as trustworthy but the evidence adduced to prove the merits of the products may be judged unsound or insufficient.

Long-term recall is, of course, important. A disadvantage of assessing recall at the stand is that this gives an indication only of the maximum number who will be able to recall something from an exhibit later. If recall is tested at the time, it is not satisfactory to try to do repeat tests later on the same groups, as Bloomberg (1929) did. If memory is not tested at the show, but the addresses of the farmers are obtained, a survey could be carried out to test long-term recall among the farmers concerned.

A commercial example has been taken here and it may be queried if such a study would be typical. In fact, these methods could be applied to almost all exhibit evaluation studies; what is unusual in this example is that the aim of the exhibit could be unambiguously defined, in terms of sales, and its success thereby made easier to assess. At museums the aim is to educate, and the investigation could be concluded at the stage of long-term learning. A difficulty arises in defining what should be learnt, for the exhibitor's ideas of what are the main points of an exhibit may well not be broad enough to include all that some visitors could learn from it. In fact, something may be learnt which appears quite irrelevant to the main purpose of the exhibit: to illustrate this, one may quote a remark made by a farmer who saw an exhibit featuring live cattle, intended to show the value of different feeding methods. The farmer came away quite happy because he had seen an exhibit which showed that 'the Ministry de-horn their cows.'

#### IV.—CONCLUSIONS.

The data discussed here make it clear that there are methods which can be used to examine the effectiveness of exhibits, and more research of this nature may allow generalizations to be made about the value of exhibitions, and about ways for improving them.

Some who use exhibits seem to despair of their value, but often without much evidence for this one way or the other. At least, exhibits deserve to be studied more fully, if only so that their effectiveness can be compared with that of other ways for transmitting information. The relative cost of exhibits as compared with other sources of communication should also be considered.

ACKNOWLEDGMENTS.—The writer wishes to acknowledge the help given to him by Miss M. Smith, Miss S. D. Carrinci, Mr. C. S. O'D. Scott, and Mr. H. D. Willcock, with whom he has often discussed this topic.

#### V.—REFERENCES.

- BELSON, W. A. (1959). A technique for studying the effects of a television broadcast. *Appl. Stats.*, 5, 195-202.  
 BLOOMBERG, M. (1929). An experiment in museum instruction. *Amer. Assoc. Museums Publ.*, New Series No. 8.



- BROOKS, J. A. M., and VERNON, P. E. (1956). A study of children's interests and comprehension at a science museum. *Brit. J. Psychol.*, 47, 175-182.
- FOOD AND AGRICULTURE TECHNICAL INFORMATION SERVICE (1958). Agricultural exhibits in advisory work. *Supplement to F.A.T.I.S. Review*.
- KNEE, A. W. (1953). On getting knowledge over to farmers. *Agriculture, Lond.*, 60, 324-328.
- LANER, S., and VERNON, M. D. (1957). Visual aids in technical training. *J. Educ.*, 89, 424-426.
- MARSTON, L. R. (1925). The emotions of young children. An experimental study in introversion and extroversion. *Univ. Ia Stud. Child Welf.*, No. 3.
- NEDZEL, L. N. (1952). *The Motivation and Education of the General Public through Museum Experiences*. Ph.D. dissertation. University of Chicago.
- NIELSEN, L. C. (1946). A technique for studying the behaviour of museum visitors. *J. educ. Psychol.*, 37, 103-110.
- ORLANSKY, H. (1951). "Parliament Past and Present" Exhibition at Westminster Palace, 1951. *Social Survey Report*, No. 188.
- ROBINSON, E. S. (1928). The behaviour of the museum visitor. *Amer. Assoc. Museums Publ.*, New Series, No. 5.
- SHEPPARD, D. (1959). The evaluation of exhibitions. *Conferences and Exhibitions*, No. 12, 45-48.
- SHEPPARD, D., SMITH, M., and CARRINCI, S. D. (1960). The effectiveness of the Ministry of Agriculture, Fisheries and Food's educational exhibits at agricultural shows. In process of publication in *Agric. Progr.*
- WILLCOCK, H. D. (1949a). "On Our Way" Exhibition at Marble Arch. *Social Survey Report*, N.S. 140.
- WILLCOCK, H. D. (1949b). The Touring "On Our Way" Exhibition at Dudley and Neath. *Social Survey Report*, N.S. 145.
- WILLCOCK, H. D. (1951). The audience at a productivity exhibition in Birmingham. *Social Survey Report*, S.S. 173.
- VERNON, M. D. (1950). The visual presentation of factual data. *Brit. J. Educ. Psychol.*, 20, 174-185.
- VERNON, M. D. (1951). Learning and understanding. *Quart. J. exp. Psychol.*, 3, 19-23.
- VERNON, M. D. (1952). The use and value of graphical methods of presenting quantitative data. *Occup. Psychol.*, 26, 22-34.
- VERNON, P. E. (1946). An experiment on the value of film and film-strip in the instruction of adults. *Brit. J. educ. Psychol.*, 16, 149-162.

(Manuscript received, Sept. 2nd, 1959.)

# SUMMARY OF RESEARCH REPORTED IN DEGREE THESIS

## A Study of the Nature of Spatial Ability and its Relationship to Attainment in Geography.

By C. COLIN TAYLOR

(Abstract of thesis submitted for the degree of M Ed., University of Durham, Oct., 1959.)

### I.—AIMS OF THE INVESTIGATION.

(a) To investigate the relationship between spatial ability ('k' factor) and (i) the ability to perceive and reproduce the proportions of a figure and (ii) the ability to perceive and reproduce details.

(b) To investigate whether spatial ability ('k' factor) is related to attainment in geography.

The first part of the study originated in a suggestion made by Macfarlane Smith<sup>2</sup> that spatial tests which depend for success on the perception of the form or structure of a figure (proportion) tend to have higher k-loadings than those which depend on the perception of details. The second part was intended to be an experimental examination of claims made by previous workers that spatial ability is important for success in geography.

### II.—THE TESTS USED.

The following twelve tests of ability or attainment were administered :

Test 1—Moray House Verbal Reasoning Test (Series 51 '52).

Test 2—Non-Verbal Test 3 (N.F.E.R., Calvert).

Test 3—Spatial Test I (N.F.E.R., Smith. Total score in sub-tests 2, 3, 4, 5, involving little imaginative manipulation of figures).

Test 4—Spatial Test I (Total score in sub-tests 1 and 6. These sub-tests involve much imaginative manipulation).

Test 5—Clerical Test 1 (N.F.E.R., Richards. Total score in sub-tests 1, 2 and 4, involving perception of detail).

Test 6—Score in Geography, obtained by standardising the marks in school internal examinations to a common mean and standard deviation.

Test 7—Objective test of locational geography (in map form).

Test 8—Objective test of geographical knowledge and reasoning (in verbal form).

Test 9—Objective test of geographical distributions (in map form).

Test 10—N.I.I.P. Test of Memory for Designs. Marked according to the original marking scheme.

Test 11—Score in Test 10 combined with a Drawing of a Man Test.<sup>1</sup> Both marked for correctness of detail, i.e., the number of items or parts which were perceived as making up the composite whole. This was intended to test an ability to perceive details.

Test 12—Score in Test 10 combined with a Drawing of a Man Test.<sup>1</sup> Both marked for correctness of proportions. The perception of proportions was to be regarded as the perception of the parts of the design in relation to the whole, i.e., a due relation in size between the parts and the whole had to be perceived.

The marking for correctness of proportions was carefully carried out by the writer, the measurements being made to one-tenth of an inch.

Tests 7, 8 and 9 were devised by the writer personally.

### III.—ADMINISTRATION OF THE TESTS.

The test population consisted of 100 third-year girls and 100 third year boys, both groups having an average age of 14 years 4 months. These pupils attended secondary commercial schools in Newcastle-upon-Tyne. Reliability coefficients were calculated for all the tests (except where given in the manuals).

These coefficients were usually high, seldom falling below .8, except in the case of the school internal examinations, for which the reliability coefficients were .457 and .379, for boys and girls respectively.

Product-moment coefficients of correlation were calculated between all pairs of variables, separate matrices being calculated for boys and girls.

#### IV.—RESULTS.

(a) *Sex-differences.* Mean scores for boys and girls were calculated for all variables, and critical ratios were determined to assess the significances of the sex-differences. These are shown in Table 1. Marked sex-differences (C.R. > 2.58) occur for all the spatial tests (3, 4, 10 and 12) in favour of boys and for the verbal test (1) in favour of girls.

TABLE 1

THE SEX DIFFERENCES FOR THE TEST BATTERY (+ = FAVOURS BOYS).

Test	Sex	Mean Score	Difference of Means	Standard Deviations	Critical Ratios
1	Boys	71.17	-5.94	8.900	4.752
	Girls	77.11		8.770	
2	B	71.55	-1.80	12.381	1.133
	G	73.55		9.951	
3	B	71.93	+10.75	13.240	5.984
	G	61.18		12.143	
4	B	26.14	+4.48	7.475	4.956
	G	21.66		6.255	
5	B	67.61	-4.04	14.748	1.865
	G	71.65		15.861	
6	B	49.73	-0.26	9.815	0.1876
	G	49.99		9.884	
7	B	49.52	-3.54	20.530	1.419
	G	53.06		14.187	
8	B	31.22	-0.74	7.669	0.724
	G	31.96		6.757	
9	B	15.52	+0.75	5.683	1.008
	G	14.77		4.808	
10	B	33.04	+3.12	6.953	3.477
	G	29.92		5.669	
11	B	63.33	-0.67	15.534	0.324
	G	64.00		13.670	
12	B	37.59	+11.90	12.073	8.005
	G	25.69		8.674	

(b) *Factor Analyses.* Centroid factor analyses were carried out on the correlation matrices for both boys and girls (cf. Tables 2 and 3), the resulting unrotated factors being interpreted in the light of the hierarchical theory of the structure of mental abilities.<sup>3</sup>

TABLE 2

UNROTATED CENTROID FACTORS (AFTER ONE ITERATION)—CENTROID FACTOR MATRIX FOR 100 BOYS.  
(with proportions of variances contributed by the centroid factors, obtained communalities and guessed communalities.)

Test	I	II	III	I <sup>2</sup>	II <sup>2</sup>	III <sup>2</sup>	h <sup>2</sup> (obt'd)	h <sup>2</sup> (guessed)
1	+·4069	+·1523	-·2762	·1656	·0232	·0763	·2651	·4038
2	+·6272	+·1056	-·1114	·3934	·0112	·0124	·4170	·5301
3	+·7631	+·4455	-·3462	·5823	·1985	·1199	·9007	·9390
4	+·7144	+·4300	-·3732	·5104	·1849	·1393	·8346	·9390
5	+·4830	-·0742	+·0996	·2333	·0055	·0099	·2487	·4135
6	+·5121	-·2152	-·1105	·2622	·0463	·0122	·3207	·4558
7	+·7064	-·5815	+·0387	·4990	·3381	·0015	·8386	·8405
8	+·6102	-·6153	-·1174	·3723	·3786	·0138	·7647	·8405
9	+·6901	-·5288	+·1171	·4762	·2796	·0137	·7695	·8035
10	+·7926	+·2861	+·3954	·6282	·0819	·1563	·8664	·8766
11	+·5289	+·2771	+·3297	·2797	·0768	·1087	·4652	·5876
12	+·7442	+·2997	+·4065	·5538	·0898	·1052	·8088	·8766
Variance {				$\Sigma a^2$	4·9564	1·7144	·8292	7·5000
				%	66·09	22·85	11·06	100%
				% Av.	41·30	14·29	6·91	62·5%

TABLE 3

UNROTATED CENTROID FACTORS (AFTER ONE ITERATION)—CENTROID FACTOR MATRIX FOR 100 GIRLS  
(with the proportions of variances contributed by the centroid factors, obtained and guessed communalities).

Test	I	II	III	I <sup>2</sup>	II <sup>2</sup>	III <sup>2</sup>	h <sup>2</sup> (obt'd)	h <sup>2</sup> (guessed)
1	+·4751	+·1790	-·4566	·2257	·0320	·2085	·4662	·4694
2	+·7298	+·2502	-·1692	·5326	·0620	·0286	·6238	·6629
3	+·7534	+·4163	+·1832	·5676	·1733	·0336	·7745	·8913
4	+·6785	+·3817	+·1837	·4604	·1457	·0337	·6398	·8913
5	+·3046	+·0913	-·6249	·0928	·0083	·3905	·4916	·4573
6	+·4844	-·4875	-·1667	·2346	·2377	·0278	·5001	·5777
7	+·6014	-·6393	+·0351	·3617	·4087	·0012	·7716	·7464
8	+·4995	-·6532	+·0819	·2495	·4267	·0067	·6829	·7464
9	+·3359	-·4663	+·0240	·1128	·2174	·0006	·3308	·4938
10	+·6949	+·3783	+·2218	·4829	·1431	·0492	·6752	·7289
11	+·3512	+·2682	+·3734	·1233	·0719	·1394	·3346	·5439
12	+·6110	+·2929	+·2556	·3733	·0858	·0653	·5244	·7289
Variance {				$\Sigma a^2$	3·8172	2·0132	·9851	6·8155
				%	56·00	29·55	14·45	100%
				% Av.	31·81	16·78	8·21	56·80%



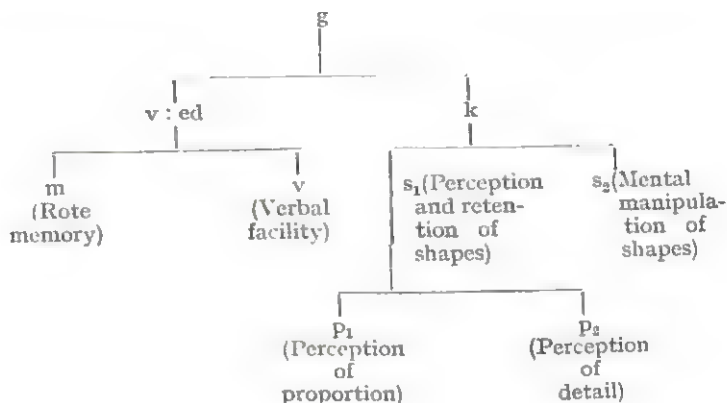
The psychological interpretation was confirmed by carrying out a group factor analysis, which gave more satisfactory results for the boys than for the girls.

The following scheme was tentatively proposed to account for the groupings of the factor loadings :

Factor I

Factor II

Factor III



## V.—CONCLUSIONS.

### (a) *The Nature of Spatial Ability ('k' factor).*

(i) Test 3 (Sub-tests 2, 3, 4, and 5 of Spatial Test I) and Test 4 (Sub-tests 1 and 6) showed marked sex-differences in favour of boys, but the most striking sex-differences in favour of boys was shown by the Memory for Designs Test, when marked for correctness of proportions (C.R.=8.0).

A small but insignificant sex-difference in favour of girls was shown by the Memory for Designs Test, when marked for accuracy of detail. Thus, marking this test for detail has produced a large 'swing' away from the normal sex-difference in spatial tests in favour of boys.

(ii) The factor analyses show that the three ways of marking the Memory for Designs Test all give rise to high spatial loadings. They also indicate that, both for boys and girls, the Memory for Designs Test has a higher 'k' loading when marked for correctness of proportions than when marked for accuracy of detail (thus confirming one of the hypotheses being investigated).

(iii) Evidence was obtained suggesting that a sub-factor  $S_1$ , involving the perception and retention of impressions of shape, contrasts with a sub-factor  $S_2$ , involving an ability to manipulate shapes mentally, both sub-factors entering into spatial ability. In the case of boys, the loading for  $S_1$  is highest in the Memory for Designs Test marked for correctness of proportions, while for girls it is highest when this test is marked for accuracy of details.

(iv) The analyses also suggest that boys are superior to girls with regard to sub-factor  $S_1$  and that girls apparently possess greater ability than boys in tests involving sub-factor  $S_2$ .

(v) The results are consistent with the view that spatial ability ('k' factor) involves a capacity to perceive and retain mentally an impression of the structure or form of a shape or pattern *as a whole* (cf.<sup>2</sup>). Tests designed to emphasise this capacity show the largest sex-difference in favour of boys.

(vi) The most striking result of marking the Memory for Designs Test for accuracy of detail is a substantial reduction in the first-factor loading ('g'). This method of marking also caused the sex-difference to 'swing' in favour of girls.

(b) *The relationship between spatial ability and attainment in geography.*

(i) There were no significant sex-differences in test performance in any of the geography tests.

(ii) The factor analyses provided no evidence that spatial ability ('k' factor) enters into attainment in geography, as measured by the school internal examination, or by the objective tests specially prepared by the writer. The loadings in the first bipolar factor indicate that scores in the examination and tests in geography are sharply contrasted with scores in the spatial tests.

(iii) The four measures of attainment in geography had higher 'g' loadings for boys than for girls.

(iv) Attainment in geography was highly dependent on verbal-educational ability and this dependence was even more marked for girls than for boys.

#### VI.—REFERENCES.

- <sup>1</sup> GOODENOUGH, F. (1926). *The Measurement of Intelligence by Drawings*. New York : World Book Coy.
- <sup>2</sup> SMITH, I. MACFARLANE (1954). The development of a spatial test. *Durham Res. Rev.* 1, 5, 19-33.
- <sup>3</sup> VERNON, P. E. (1950). *The Structure of Human Abilities*. London : Methuen.

## NOTES AND COMMENTS

### COMMENTS ON ARTICLES BY MR. D. G. LEWIS AND MR. W. R. JONES

By J. L. WILLIAMS

(University College of Wales, Aberystwyth)

Reports of attempts by Mr. D. G. Lewis and Mr. W. R. Jones at comparing the attainments of bilingual and monoglot children in Wales may be of some academic interest and their disagreements can be regarded as somewhat amusing, but it is very doubtful whether investigations of this nature can ever make a worth while contribution either to the study of bilingualism or to the means of attaining a high degree of bilingualism among communities where such an educational aim has been accepted. Mr. W. R. Jones' reference to "somewhat contradictory results" already echoes through very many investigations in many countries, and more variations on the same theme can hardly be expected to be anything but tiresome.

The findings of both writers are bound to be inconclusive and of little value as long as they continue along their present lines of investigation. It appears that no significant step forward has been taken in the planning of such investigations since the pioneer work done by Saer in 1922 and 1923. Any advance made has been limited to the use of better tools and improved statistical methods.

Investigators have consistently failed to isolate bilingualism as a variable. Even the importance of socio-economic factors has been ignored until Mr. Jones made a belated 'discovery' of its significance in his latest publication.\* This is strange in view of the fact that it has been known for many years that the scores of groups of children in various tests tend to vary in accordance with such factors as parental occupation and other background variables. In fairness to the two investigators and their predecessors in Wales, one should, however, note that many other research workers in the same field elsewhere have unwittingly committed similar sins of omission, with the result that their investigations too have failed to produce useful knowledge, let alone valid conclusions, concerning the effect of bilingualism *per se* upon child development and educational progress.

Such examples of *non sequitur* as Mr. Lewis's statement, "One can conclude then that linguistic background can be regarded as an important factor of the school differences in attainment in English which are not explained by differences in intelligence" are common in reports of investigations of this nature. Is it not possible that several schools "drawing most of their pupils from Welsh-speaking homes," in his sample, were carrying out the language policy recommended by the Welsh Department of the Ministry of Education and, consequently, did not make any use of English in the education of pupils under the age of seven, and even in the junior school devoted much time to the development of command of the mother tongue (Welsh)? If schools in his sample followed what is the normal practice in Welsh-speaking areas, it would be much more reasonable to conclude that school differences in attainment in English and arithmetic arose from school language policy, and in particular from variations in teaching time and language contact under school conditions, rather than from differences in linguistic background outside the schools, in spite of the evidence provided by statistical work done by Mr. Lewis, Mr. Jones, and others.

It is suggested respectfully, that the time has come for potential research workers who are anxious to tackle problems associated with bilingualism in Wales to be directed into more fruitful soil than that to be found in the often tilled, but most unproductive field in which intelligence, school attainments in English, etc., at junior school level only have so often been compared without any consideration being given to variations in teaching time, methods of language presentation and usage, number of subjects taught, and to important variations in home background not, of necessity, linked with language spoken or even with parental occupation.

\* *Bilingualism and Intelligence*, Cardiff, 1959.

## REPLIES TO COMMENTS BY J. L. WILLIAMS

Dr. J. L. Williams objects to my conclusion that "linguistic background can be regarded as an important factor of the school differences in attainment in English which are not explained by intelligence," on the grounds that some, but not all, schools were carrying out a special language policy, a policy which could well account for the differences in test results. This possibility, however, had been considered from the very outset. The head teachers of all the schools were asked about their language policy, and a detailed statement was received from each school. No marked differences in the schools' language policy were found to exist. In all schools, the medium of instruction (of the subjects tested) was English, Welsh being used to but a very small extent. All schools, too, devoted separate teaching periods to the development of Welsh. Minor variations in "language contact under school conditions" must be viewed against this general uniformity of practice.

In an earlier paper on bilingualism and non-verbal intelligence\* groups selected on the basis of linguistic background were compared in test performance. Each group contained children from all the schools sampled. Note was taken of relevant non-linguistic differences between the groups. Thus, it was found that the groups did not differ appreciably in age; and that the ratio of boys to girls in each group was practically the same; moreover, since each group was spread over all the schools, such internal school differences that may coexist with a common language policy (e.g., variations in subject time, 'formal' *versus* 'liberal' quality of teaching, etc.) would tend to appear solely as *within-group* differences. Careful consideration was given to group differences in socio-economic status (despite the 'amusing' and ill-founded criticisms of Mr. W. R. Jones), and also in urban-rural environment.

Of course, the results of this and other comparative studies—and there are many carefully conducted researches with similar results from outside Wales to be assessed—do not *prove* bilingualism to be a source of intellectual disadvantage during childhood. Neither does a significantly higher incidence of lung cancer among heavy smokers prove a causal relationship. But the proper function of the critic is to draw attention to specific factors which have been overlooked. Unsubstantiated attack and polemical over-statement have little to contribute.

D. G. LEWIS.

I should like to comment briefly on Dr. Jac L. Williams's references to my recent article on 'Bilingualism and Non-Verbal Intelligence':

1.—My criticisms of Mr. D. G. Lewis's article were mainly concerned with (a) a matter of principle, namely the necessity of acknowledging clearly and fully the source of one's experimental data, and (b) with the importance of an adequate assessment of socio-economic as well as linguistic factors in any experimental study of bilingualism. Far from being 'somewhat amusing' disagreements of an academic nature, as Dr. Jac. L. Williams avers, these criticisms seem to me to be of practical importance in relation to the future progress of educational research in Wales.

2.—Furthermore, Dr. Williams shows considerable temerity in asserting that no significant step forward has been taken in the planning of 'bilingual' investigations since the pioneer work done by Saer in 1922 and 1923. Only a person who possesses a general rather than a specialised knowledge of psychological studies in the field of bilingualism would be rash enough to make such a bold assertion. A detailed study of the investigations of Barke and Parry-Williams, undertaken at Cardiff, in the 1930's and of my own researches undertaken at Cardiff and Bangor, from 1933 to the date, as well as the investigations of some of his own colleagues and post-graduate students at Aberystwyth, will show beyond doubt that real progress has been achieved, particularly in the use of analysis of covariance as an integral part of experimental designs, in the assessment of the influence of reading and linguistic factors on test-scores, and in the application of attitude tests and language questionnaires in the study of second-language learning and teaching. It will also be evident that as a rule every possible care has been taken in recent surveys to isolate bilingualism as an experimental factor. It may very well be that this isolation has been only

\* This *Journal*, XXIX, 17-22.



partially successful, but, as every experienced worker in the field of educational research will know, complete isolation of any particular variable is often very difficult to achieve, especially so in experiments which are complicated by a multiplicity of interacting human and social factors.

3.—Dr. Jac L. Williams's complaint that no consideration has been given to variations in teaching time, methods of language presentation, etc., in comparative studies of monoglot and bilingual children will not bear examination for one single moment. I would strongly advise him to read *in detail* my own researches which have been published from time to time as articles in this *Journal* and as a series of monographs by the University of Wales Press Board, so that he may see for himself that considerations of this kind were consistently borne in mind in the process of interpreting test results.

4.—Finally, I would point out that, as far as Wales is concerned, the scientific study of bilingualism from 1922 onwards has been pursued by one or two researchers as a rule working alone, whereas others, although equally well placed to undertake researches of a similar kind, have been content to do nothing other than indulge in uninformed and irresponsible criticism. I would respectfully suggest that the time has now come for Dr. Jac L. Williams and other 'potential' researchers to cease to pontificate generally on the topic of bilingualism and to become active participants and practical workers in this most difficult field of enquiry. W. R. JONES.

## ON D. H. STOTT'S ARTICLE: INTERACTION OF HEREDITY AND INTELLIGENCE IN REGARD TO 'MEASURED INTELLIGENCE'

Educational psychologists will be grateful to Dr. Stott for raising afresh the far-reaching problem of environment *versus* heredity, and for his interesting summary of recent work in experimental biology. Unfortunately, however, his survey is limited almost exclusively to those publications which support, or seem to support, his side of the case; and, as readers familiar with his earlier writings will know, where intelligence testing is concerned he has strong and somewhat unconventional views. "The legend that the intelligence test measures innate unalterable ability," he has told us, "was fathered by an earlier generation of psychologists fighting for recognition, and mothered by administrators glad for someone else to carry the baby . . . Any clinic which abandoned intelligence tests entirely would not only be able to handle more cases, but increase its success rate many times."\*

As a matter of history, intelligence tests were fathered by Sir Francis Galton and Alfred Binet (Professor at the Sorbonne) comparatively late in life. Neither had then to fight for recognition. In this country the earliest administrators who encouraged the use of such tests were the education officials of the London County Council; and in the first memorandum which they issued on the subject (a revised version of Binet's tests for use of teachers and doctors diagnosing defective children) stress was laid on "one fundamental truth: tests, far more scientific than those here set out, can still be but the beginning of the examination, never the end, of the child,"† In varying terms all of us in those days emphasized that 'intelligence,' in the sense of *innate* general ability, was a hypothetical factor, whose amount could at best only be estimated, and that the marks or measurements obtained with tests formed merely one of the many aids to making such an estimate. Some writers like to distinguish the two concepts by using the labels 'intelligence A' ('innate') and 'intelligence B' (measurable). Dr. Stott uses the term 'intelligence' to cover both, often without indicating which he has in mind. But no psychologist of repute has ever supposed that a crude test-measurement, taken just as it stands, provides an infallible assessment of 'innate unalterable ability.'

Dr. Stott, however, appears to hold that the whole notion of 'general ability' is an untenable hypothesis. "Our ability to measure *specified* differences," he says,

\* D. H. STOTT (1956). *Unsettled Children*. pp. 27-8.

† C. BURT (1921). *Mental and Scholastic Tests*. p. xv.

"does not justify us in assuming that there is a unitary factor underlying them." Certainly not: the justification lies in the positive correlations (usually quite high) between such measurements—correlations which practically every investigator has confirmed. And the statistical inference that these are due to a concrete 'factor' is borne out by a vast array of evidence\* from other fields—experimental, introspective, biological, and above all, neurological—evidence which Dr. Stott does not even mention, let alone refute.

But what Dr. Stott is chiefly concerned to dispute is the idea that differences in 'measured intelligence' are in any appreciable degree due to genetic influences. He begins by stating that the attempt to distinguish hereditary influences from environmental "has been called in question owing to the demonstration of a certain correlation," between the two. Maddox, he says, has pointed out that "this correlation is due to the sociological fact" that high intelligence in the parent is associated with cultural stimulation in the child. Maddox did not point out this causal process; he merely thought that such conditions *might* account for *all* the correlation between parent and child—a purely speculative suggestion. Dr. Stott apparently assumes that, when two such factors are correlated or interact, it is impossible to assess by statistical analysis "the relative contributions" of each. But Sir Ronald Fisher long ago showed how, by an analysis of variance, this could be done, and applied his techniques to the problem of heredity and environment. If Dr. Stott believes that Fisher's demonstration is faulty, he should at least explain where he supposes Fisher has slipped up.

Dr. Stott goes on to cite a series of experimental studies indicating that congenital malformations in animals may be due partly or wholly to "environmental stress during gestation." Vitamin deficiency, for example, is liable to produce hernia in certain strains of rats known to inherit a predisposition of this abnormality. Similar malformations result from a wide variety of disturbances in pregnant animals—notably among mice, pigs, guinea-pigs, and fowl. In man therefore (so he supposes) analogous processes may occur, and bring with them a defect or a diminution in the child's intelligence.

This hypothesis he describes as a "new biological concept." But, so far as mental defect is concerned, it has been the subject of discussion and research for half a century or more. To-day no one would deny that in certain cases an unfavourable uterine environment may contribute to the production of certain types of mental deficiency, often marked by characteristic physical symptoms. But equally, I imagine, most authorities are now agreed that such cases are comparatively rare. Most types of malformation, even if they affect the nervous system, are *not* associated with mental defect; when they are, the defect is usually grave enough to amount to definite idiocy, and nine out of ten of the victims die during the first few years. The two instances cited by Dr. Stott are those of anencephaly and mongolism. Cases of anencephaly (a defect which is far more extensive than mere "lack of cerebral cortex") rarely survive for more than a few days. As for mongolism, Dr. Stott contends that, unless we recognize that "pre-natal environment plays a large part, it would be hard to account for" its frequency in children of elderly mothers. Nevertheless, within the last few years, it has been amply shown that this particular defect results from extra chromosomal material—a genetic change which takes place at or before conception, and can have nothing to do with the uterine environment, as was formerly believed.

Dr. Stott applies the same hypothesis to account for the high correlations between estimates of intelligence in the case of twins, even when brought up in widely different post-natal environments—as in the studies reported by Miss Conway and myself. Since twins are carried in the same uterus at the same time, the uterine conditions, he maintains, must be the same for both; and, since in human beings, the presence of two foetuses in the womb is so unusual as to increase 'pregnancy-stress,' the effect will be to produce similar defects, mental as well as physical, in *both* members of the pair. The defect will vary in severity with the severity of the

\* See: The evidence for the concept of intelligence, this *Journal*, XXV, pp. 158-177.

stress. Hence, the natural result, he argues, must be an increased correlation between the intelligence of twins.

But, as a matter of fact, in 'non-identical' twins the correlation differs hardly at all from the correlation for ordinary siblings (see this *Journal*, XXV, p. 168, Table 1). So here there is no indication of the causal process alleged. Dr. Stott, however, if I understand him rightly, assumes that there must also be some pre-existing genetic tendency which the pre-natal stress 'facilitates' and that this is found in 'identical' (i.e., one egg) twins. Yet, if that is his explanation, then he is tacitly admitting that the genetic predisposition is the decisive factor.

However, as numerous embryologists have pointed out, the uterine conditions are *not* the same for both members, since they cannot both occupy the same position. Usually, the second to be delivered (often still born) is the one that has suffered most. The fact that 'identical' twins develop within the same membrane as well as within the same uterus, would appear, if anything, only to aggravate the difference.

But do such conditions "make twins more similar in mentality" or lead to "the production of definite defect?" This after all is a question of fact on which there is plenty of data. In inquiries carried out in I.C.C. institutions and reported by Miss Conway, twenty-six cases of mental deficiency in an 'identical' twin were discovered, and in twenty-four of the cases the other twin was *not* defective. In these institutional cases the deficiency was of a low grade pathological type, usually accompanied by physical symptoms. Outside institutions, where the deficiency is as a rule merely an excessive deviation below the general average, of a kind fitting in with the normal distribution, the correlation with physical signs was negligible.

These findings are fully in keeping with those reported by most other investigators. For example, in a recent survey, Berg and Kirman have found that, in forty-nine cases of mental deficiency occurring in a twin, the other twin was defective in only four cases.\* Thus, so far as intelligence is concerned, the type of causation invoked in Dr. Stott's rather speculative arguments proves to be the exception rather than the rule. Certainly, the meagre amount of evidence that he himself adduces can by no means suffice to overthrow a conclusion accepted by such experienced geneticists as Fisher, Mather, and Darlington, and indeed, if I am not mistaken, by almost every leading authority in this country.

CYRIL BURT.

## REPLY

Sir Cyril Burt must have misunderstood the theme of my article. I did not raise the problem of environment *versus* (repeat *versus*) heredity. This is precisely the dichotomous approach I tried to show was unreal. We are dealing with complementary conditions for development. Up to a certain point heredity-environment is a useful way of conceptualizing two main classes of necessary conditions. But to polarize them into a dichotomy is epistemologically primitive. Equally primitive is the vitalism of calling them 'influences,' making them conceptual guinea-pigs upon whom to pin a percentage responsibility. Nor am I, naturally, "concerned to dispute . . . that differences in 'measured intelligence' are in any appreciable degree due to genetic-influences." I just do not take sides in this unreal battle. Sir Cyril will have to look elsewhere for an environmentalist to serve him as an opponent.

I am surprised that he still regards positive test-correlations as implying a concrete factor of general ability. Has it not been pointed out many times that the general factor which emerges depends upon how the data fed into the machine are selected? If the psychologist arranges the test-material in accordance with a concept of 'intelligence' dominant within his culture-pattern he will obtain a general factor consistent with that concept, and, of course, he will identify it there-

\* BERG, J. M., and KIRMAN, B. H. (1960). "The mentally defective twin." *Brit. Med. J.*, June 25th, pp. 1911-17.

with. The 'general factors'—and, indeed, much that we ordinarily regard as evidence of 'intelligence'—could be equally seen as a function of motivation and at one remove, of the mental skills which are also a function of motivation.

I disagree also with Sir Cyril Burt's time scale. Foetal damage by environmental insult was demonstrated in a mammal only in 1937 by Hale and in man by Gregg in 1941. A definite genetic-environmental interaction was demonstrated only in 1949 by Andersen and in 1954 in more detail by Fraser. In so far as the earlier geneticists used the term 'facilitation' it was in quite a different sense, namely, the facilitation of the action of one gene by another gene.

I would not agree that unfavourable uterine environment is only rarely responsible for mental defect, and have in fact produced evidence (1957, 1959) to show that some 70 per cent. of mental subnormality in normal-appearing children may be thus caused.

As regards mongolism, again the time-scale: Turpins actually published the first report of an extra chromosome in 1959. It has not been "amply shown that this particular defect results from extra chromosomal material." Results to date, while indisputable, are fragmentary: we do not yet know whether mongols are mosaics, nor the chromosomal constitution of their parents, nor just when or how the genetic anomaly arises. With little doubt anomalous genetic constitution is one of the pre-conditions for mongolism, but it is epistemologically naive to jump to the conclusion that is it the cause. It must not be forgotten that a large proportion (Hertig) of early embryos are malformed, probably because of genetic error. Normally, these do not develop to term, but in the case of mongols and other genetically anomalous malformations the normal mechanism for their being weeded out seems to fail. I have awaiting publication evidence that this is due to maternal stress (hormonal disturbance?) predominantly at the second and third months of pregnancy.

D. H. STOTT.

#### MAIN REFERENCES

- GREGG, N. McA. (1941). Congenital cataract following German measles in the mother. *Trans. Ophthalm. Soc. Austral.*, 3, 35.  
 ANDERSEN, D. H. (1949). Effect of diet during pregnancy upon the incidence of congenital hereditary diaphragmatic hernia in the rat. *Amer. J. Path.*, 25, 163-84.  
 STOTT, D. H. (1957). Mental and physical handicaps in the child following a disturbed pregnancy. *Lancet*, 171, i, 1006.  
 STOTT, D. H. (1959). Evidence for pre-natal impairment of temperament in mentally retarded children. *Vita Humana*, 2, 125-48.



## CRITICAL NOTICE

### THE CROWTHER AND ALBEMARLE REPORTS

*Fifteen to Eighteen: Report of the Central Advisory Council for Education.*  
H.M. Stationery Office, 1959. Vol. I, pp. xxxii+519, 12s. 6d.

*The Youth Service in England and Wales: Report of the Committee Appointed by the Minister of Education.* H.M. Stationery Office, 1960, pp. iv+136, 6s.

By CYRIL BURT

Four years ago the Minister of Education invited the Central Advisory Council, of which Sir Geoffrey Crowther is chairman, "to consider, in relation to the changing social and industrial needs of our society and the needs of its individual citizens, the education of boys and girls between the ages of 15 and 18." Their Report presents a critical survey of the present system, and devotes special attention to the secondary modern schools, the sixth forms in the grammar schools, and the various types of further education at present available. The inadequacies which their investigations have disclosed demand, so the Council contends, a methodically planned educational programme covering the next twenty years.

Their essential recommendations may be summed up as follows: (i) a compulsory extension of the school leaving age to 16 by some date between 1966 to 1969; (ii) a series of experiments which will lead to a scheme of compulsory part-time education, to be instituted during the 1970's, for all youths and girls of 16 and 17 who are not then receiving full-time education; (iii) further experiments, leading to the introduction of external examinations below the level of the G.C.E., organized on a regional rather than national basis, for pupils in modern schools during their fifth year; (iv) immediate measures to forestall premature or excessive specialization in grammar schools; (v) a closer integration of the training given in school with that provided by the schemes of further education, based largely on a more coherent national system of education in the widest sense; (vi) the substitution of two school leaving dates for three; and (vii) last but by no means least, more effective steps to meet the urgent need for improving both the quantity and the quality of teachers. The total cost of all these recommendations (estimated by adding the amount that will be lost in production to the amount directly entailed by the new provisions) is put at £200-250 million a year. But, it is argued, this rather heavy burden must be regarded as "a profitable investment in national efficiency." More than half the extra expenditure will be the effect of raising the school leaving age; and it is with this particular item in the Report that most of the discussions in the popular press have been concerned.

Some of the most interesting findings of the Council are contained in a second and supplementary volume. They include the results of three factual surveys, sponsored by the Council, namely (a) a social survey, dealing more particularly with the home background and employment record of school leavers and with factors affecting the age of leaving school; (b) a national service survey, dealing with the distribution of latent ability, and with the occupations, earnings, and leisure activities of recruits, and (c) a survey of technical courses. The Report itself starts with the question: "Who gets full-time education?" The figures obtained from recruits for the services indicate that 72 per cent. of them left school at 15 or earlier; however, about two-fifths of those aged 15 to 17 got some kind of further education, if only for one day a week. But what to the psychologist will seem especially regrettable is the number of intelligent youths and girls who had failed to secure any kind of higher education commensurate with their capacities.

In their analysis the Council adopt the "Army measure of 'ability groups'," based on a battery of five intelligence tests and yielding a conventional classification into six groups or grades: namely (if we convert the percentages to terms of an I.Q. scale with an S.D. of 15): (i) high ability (the top 10 per cent. i.e., over 120

I.Q.); (ii) moderately high ability (the next 20 per cent., i.e., over 108 I.Q.); (iii) good average (100 to 108 I.Q.); (iv) poor average (92 to 100 I.Q.); (v) moderately low (under 92 I.Q.); (vi) very low (under 80 I.Q.). Of the men in the highest group 9 per cent. left school at 15 and 33 per cent. at 16; while 58 per cent. stayed on still longer, usually leaving at 18. But most of the latter were drawn from the professional, managerial, and clerical classes. Among the sons of the manual workers, 19 per cent. of those in the highest ability group left at 15 and 44 per cent. at 16, while only a quarter stayed on till 18. Of the men in the next highest ability group 65 per cent. left at 15, 22 per cent. at 16, and only 7 per cent. stayed on till 18. All this seems at first sight to imply a gross wastage of talent; and it is largely to rectify this loss that the Report recommends the compulsory raising of the school leaving age to 16.

The reasons advanced in support of this proposal are threefold: (i) "Leaving at 15 is so deeply embedded in certain parts of the social structure that it is unlikely that this waste of talent can be remedied without compulsion."; (ii) "The demand for more educated workers is growing at almost every level of industry, and raising the age to 16 would give those near the bottom a better foundation"; (iii) "The strongest part of the case is the need for a secondary education for all; which will extend through the important and difficult period of adolescence."

The suggestion that the education of every child should be prolonged until the process of maturation is completed forms unquestionably a highly attractive ideal. But the real question is whether such an ideal can indeed be achieved by a change which is to be carried out through the schools, and is imposed by compulsion, and will, after all, add only one further year. Moreover, it is a question which turns on matters of psychological fact rather than of ethical or sociological theory. Unfortunately, at present, there is very little factual evidence available either one way or the other. The most we can do is to forget for a moment our cherished ideals and our very natural hopes, and weigh up quite impartially the somewhat slender probabilities which attach to the various pros and cons. Perhaps the only thing the psychologist can safely affirm is that children generally, and particularly children who have just passed the stage of puberty, vary so widely in so many unexpected ways that a single sweeping regulation is bound to affect different individuals very differently.

To begin with, the benefits received from any training which may be given in school must depend, not merely on the general intellectual ability of the child, but also upon his special abilities and disabilities, and, still more, on his emotional and volitional reactions. Among the after-histories of children of high ability who have been transferred to grammar schools, and yet have failed to fulfil their early promise, the commonest cause of failure is a lack of the appropriate qualities of temperament or character. Most of us who have ourselves enjoyed such an education can recall, among our own contemporaries, many individuals of an ability quite equal to our own, who left school at 14 or 15, and yet eventually achieved equal or even greater success in their own particular walk of life, in spite of, and often because of, the fact that their career took a different course. It may, therefore, be plausibly argued that some—I do not say all—who voluntarily leave school at the earliest permissible age would not really gain by another year's compulsory education, however high their actual ability.

This, I venture to think, holds still more strongly of the 'lower ability-groups.' It is here that the Crowther Report is least convincing. With children in these lower groups the processes of mental maturation and development generally come to an arrest at a much earlier age than with the average or the supernormal. They are, therefore, quite unlikely to profit by a further 12 months at school if the education given is of the prevalent type. At times, the Report speaks as though in such instances prolonged instruction might actually compensate for their innate intellectual inferiority. That certainly is a fallacy. The Council are on surer ground when they suggest that "the additional year should not be simply a continuation of what has gone before, but should offer new and challenging courses." "Can the school do this?" they ask; and their answer has a distinctly diffident ring: "We think it can, but are very far from asserting that this is yet the case." If so, the more

economical and prudent policy is surely to make tentative experiments in these new directions first of all.

Some information on the problem may, perhaps, be drawn from the experience already gained by raising the leaving age from 14 to 15. Recent investigators assure us that the extra year of education has had little effect on the educationally subnormal or on the degree of illiteracy and semi-illiteracy in the general population. The prophesied reduction in delinquency has certainly not been realized. Before the age was raised, the duller 13-year-olds, as most school teachers will recall, were far more prone to petty delinquency both in school and out of it; and when at 14 they at last gained a longed-for sense of self-dependence and came under the new discipline of industry, the number of offences declined. In 1947, however, as the Report very frankly admits, "when the school leaving age was raised to 15, there was an immediate change": the age of 14 became the "peak age for juvenile delinquency . . . and has held that place consistently ever since." But perhaps the most convincing comments come from the class teachers who will have to shoulder the burden of these duller groups. Visit the secondary modern schools in any of the larger industrial areas, and ask the masters who have had charge of the top age-group in the C streams: you will find that, almost without exception, they are highly sceptical about the advantages that are predicted. They will, moreover, put their finger on a further drawback which the Council have apparently overlooked. Not only are these unwilling youngsters quite impervious to instruction and control, but their very presence in the classroom renders the task of teaching their more willing school-fellows far more arduous and ineffectual.

This, however, is a criticism of the schools as they are organized at the present day. Hence, it is not to my mind wholly conclusive. Most of those who have discussed the proposal seem to have forgotten a relevant piece of past history. For pupils who are definitely subnormal, the compulsory extension of the school period would be by no means a complete innovation. Under the older Education Acts children who were certified as mentally deficient and transferred to special schools were obliged to attend until they were 16. Those of us who are old enough to recall the magnificent work done by the schools for elder boys and elder girls are therefore tempted to inquire whether the experience then gained might not be profitably exploited once again for the same type of child who would now be grouped with the 'educationally subnormal.' Might it not even be extended to those who are somewhat above what was formerly the borderline for certification? There is one strong point in its favour. Before the special schools were introduced, mental deficiency was commonly cited as an important cause of juvenile delinquency; by 1920 it had become, at any rate in this country, one of the least important.

However, psychologists are by profession trained to distrust armchair deductions and to demand a series of first-hand experiments before they pronounce any final verdict. Accordingly, they at any rate, should surely press for an actual trial of these new proposals on a limited and inexpensive scale, before the change is made universal. This was, in fact, the way in which the special schools I have just described were actually introduced. Under the earlier Acts, which were permissive rather than obligatory, one or two of the more enterprising authorities tried out a pilot scheme first of all. Let me add that what perhaps are most urgently needed are experimental studies of the various types of practical training that are feasible under classroom conditions to see whether any of them are likely to be really effective and, if so, which types of training suit which types of adolescent boys and girls.

Nevertheless, the school alone cannot do all. Much in the attitude of the unwilling pupils is just a second-hand reflection of the attitude adopted by his family and his friends. The Advisory Council, as we have seen, thinks that the parental objections spring chiefly from the fact that "leaving at 15 is deeply embedded in the social structure." However, "leaving at 15" is too recent a change to have become part of the working class tradition. The parents of the present generation, like the grandparents before them, left school at 14 or even earlier. In these families, therefore, both sons and daughters are assumed to 'come of age' on their fourteenth birthday, that is, directly they have reached what the lawyers call *actus plena pubertatis*, and are, thereby, transformed from 'children' to 'young persons' By



then, it is supposed, they are sufficiently mature to become financially independent of their parents, and old enough to start contributing to the family income. To-day, no doubt, most of the parents could easily afford to support their youngsters for a much longer period than was possible a generation ago; and many are already beginning to realize that the youth who has profited by a further spell of education will be far more likely to secure a good position in the industrial world than if he had left school at the earlier age. The demand for better educated employees is rapidly increasing. Already there is an urgent need, not merely for more technologists of high intelligence, but also for more technicians of moderate or average ability. And for those who look forward to such posts an extra year or so of schooling is unquestionably an asset.

But the rise in economic standards has not only sent up the incomes of the parents; it has led to an increase in the earnings of the young, which is tantamount to a domestic revolution. Many a youngster at 15½ is not only able to keep himself, but his weekly wage packet endows him with a lavish purchasing power which he is quite unfitted to use intelligently. And it is the allurements of these high earnings that induces so many of them to leave as soon as the law permits. But, it seems extremely doubtful whether a new Act of Parliament which forced such youngsters to remain for another year against their will and against the wishes of their parents would prove an effective remedy. Already, the number of those who are sensible enough to stay on has begun to increase; and, as they and their parents appreciate the advantages to be gained, and as the schools themselves remodel their curricula so as to provide a more attractive and more suitable training, it will continue to increase at an accelerated rate.

I think then we may reasonably concede that, on social grounds at any rate, the raising of the age of schooling for every pupil is highly desirable as an ideal. But, before it can safely be put into practice, far more first-hand evidence is required regarding the psychological and educational aspects of the problem. And this means further investigation. During the last ten years the total amount spent on education has risen from under £300 million to nearly £700 million; and of this only £50,000 (about 0.008 per cent.) is at present spent on educational research. The cost of preparing the Report was over £10,000. Could not a similar sum now be allotted to some competent body, such as the National Foundation for Educational Research, to carry out planned investigations and arrange the necessary experiments?

Meanwhile, there are a number of entirely non-controversial measures that ought certainly to be completed first. In far too many schools up and down the country, the higher classes are large and over-crowded; the provision for practical work is lamentably inadequate; and the number and qualifications of the teachers fall far below what is required. These defects must be remedied before any new or highly debatable proposal is put into practice. And, by way of preparing the ground, much more must be done to make the proposal more acceptable to the public, and particularly to the parents of the coming generation.

Space does not permit a detailed discussion of the remaining recommendations put forward by the Council. For the most part they involve few psychological issues; and appear to have been pretty widely accepted. Some of them—particularly the proposals for increasing the number and improving the standards of those who enter the teaching profession—are plainly far more urgent than the extension of the school period.

The problems discussed in the Report on the Youth Service overlap at many points with those discussed in the Crowther Report. The Committee was appointed in 1958 by the Minister of Education and has worked under the chairmanship of the Countess of Albemarle. As the Preface to their Report observes, it was set up at a crucial moment. During the last two or three decades many features in our national life have undergone rapid and even startling changes, which have affected the youth of the country in numerous ways requiring urgent investigation. The Youth Service itself has drifted into a critical condition, largely as a result of recent neglect. In addition, there is the coming bulge in the adolescent population; and this of itself will create new problems which have not as yet been adequately examined. For all



these reasons the Committee resolved at the outset to issue its recommendations with the utmost speed, and was therefore unable to undertake any extensive surveys or researches of its own. So far as they go, the proposals put forward will be generally welcomed. How far they can be successfully implemented, and whether they will fulfil all that the Committee hope for, are issues that time alone can decide.

The Report opens with an instructive review of the history, scope and limitations of the Youth Service, as set up in 1939, and then goes on to consider in some detail the changing pattern of adolescent life. The adolescents of to-day are taller, heavier, and more mature than their parents were at the same chronological age. During the last half century the average gain in height has been about 1-inch per decade and 1 pound in weight. Most boys and girls now reach puberty well before the age of 15. These changes, it has been suggested, may be in part the result of an underlying genetic change. During the preceding centuries there was a considerable amount of inbreeding. With the advent of the railway, and still more of the motor car, the motor bus, and the aeroplane, geographical isolation has become a thing of the past; and within the present century social barriers have been gradually broken down. As a result outbreeding has started to take the place of inbreeding. And it is well known that in both animals and plants such a change in the mating system is frequently followed by what is known as 'hybrid vigour.' However, pending further research on this and allied problems, it will probably be safer to assume that more obvious causes are chiefly responsible. Early maturation is by no means such a novel phenomenon as is commonly supposed. The age of puberty was quite as early 300 years ago: and there is ample evidence to show that, with the change to urban conditions that accompanied the industrial revolution, the rate of growth was progressively slowed down. To-day, with better food, improved social conditions, and greater opportunities for physical exercise and recreation in the fresh air, we seem now to be reverting to the more normal growth-rate of our ancestors.

To meet the altered conditions of adolescent life the Committee suggests that the Minister of Education should "initiate a ten-year development programme divided into two stages." The first should consist in what they term a period of "blood transfusion," that is, a stage of special short-term measures, supported by emergency grants, to meet the more immediate needs, particularly those of the impending bulge. The second phase, we are told, should consist in "a sustained and continuous nourishment" of the several services involved. To supervise and organize the necessary schemes a new advisory council should be appointed, to be called the Youth Service Development Council. The objects of the Service, it is emphasized, are not (as is so often supposed) "merely negative—a mere means for 'keeping young people out of trouble' or 'off the streets': rather they must be positive and constructive." The Report sums them up under three main heads: first to provide associations of the right sort; secondly, to supply a training suited to the varying needs of the young people concerned; and thirdly, to furnish opportunities for self-expression and self-development adapted to the interests and abilities of different individuals, so that both youths and girls may have a chance to display, and to respect, pre-eminence in other fields than the merely academic, instead of being content with the cruder forms of pleasure and excitement prompted by their own primitive impulses, or with the more sophisticated forms devised by those who find it profitable to exploit these primitive emotions with the aid of modern mass-media and other commercial devices.

In order to fulfil the Committee's aims, the first essential will be far more full-time leaders. These, it is argued, could best be recruited from the ranks of teachers, social workers, and "mature persons with a natural gift for leadership" and the new recruits should themselves first undergo a special course of training which will confer a professional qualification and offer a distinct career. In addition, "a generous and imaginative building programme will be indispensable." What is apparently envisaged is a centre or club, not unlike a college 'union,' with a coffee bar, a games room, a reading room, and the like, linked with various constituent self-organizing bodies which will arrange such activities as dancing, singing, amateur theatricals, camping, motor cycling, and the like, in which girls can join, as well as team-games in which they cannot.

Most of these activities will require and furnish their own type of discipline. And, it is to be hoped, throughout the Youth Service the ultimate goal will be to achieve what in the language of the old-fashioned pedagogue would be called 'internal discipline' rather than 'external discipline,' and to secure it by 'indirect methods' rather than 'direct.' In the past, the avowed educational and moral aims which formed the *raison d'être* of such clubs proved their undoing. After all how many of us would join a club whose professed object was to improve its members' characters? Even the word 'leader' has a suspicious ring. The adolescents of to-day are sheep who strongly resent a shepherd. Consequently, those who are most in need of the benefits which the youth services can provide tend to stay outside. Too often, it would seem, the Report seems to have forgotten that it is dealing with inexperienced teen-agers and Teddy-boys, who, though fully adult in body, have a mentality corresponding to a Terman-Binet age of barely 14. Half of them will never exceed the intelligence of children of 12 or 13 or even less. But, as the Committee remark at the end of their recommendations, they do not profess at this juncture to offer guidance in regard to details: "Much needs to be learnt in the light of further experience." And the difficulties we have mentioned will doubtless be closely watched by the new Development Council.

Although each of the Reports contains many minor points on which both the practical teacher and the psychological theorist will be tempted to proffer criticism, they have between them accomplished a most useful task. They have uncovered a series of urgent problems which parliaments, governments, and local bodies have so far lacked the courage, or the power, or the necessary knowledge to attack on systematic lines with the necessary financial aid; and they have brought to light a number of deeper issues of which the general public has hitherto been wholly unaware.

## BOOK REVIEWS

*Journal of Child Psychology and Psychiatry and Allied Disciplines* (ed. by C. B. Hindley, E. Irvine and E. Miller). London: Pergamon Press, Vol. 1, No. 1, Jan., 1960; to appear quarterly. Annual subscription £5 5s. (private individuals), £7 (libraries and institutions).

We welcome the launching of yet another new journal in this country which, while likely to overlap but slightly with the *British Journal of Educational Psychology*, certainly merits the attention of our readers. Its aim is to include contributions from any discipline which further the understanding of children. However, the content of the first issue suggests that it will be primarily concerned with abnormalities of childhood, five out of eight authors (or pairs of authors) being psychiatrists, and two others clinical psychologists. Perhaps one of the most valuable outcomes will be the co-operation that it brings about between the medical and psychological approaches to children.

Unfortunately, the Journal is expensive, though a considerable reduction is made to members of the Association of Child Psychology and Psychiatry. Apart from a somewhat excessive width of line, the format, type and paper are attractive.  
P.E.V.

ANDRY, R. G. (1960). *Delinquency and Parental Pathology*. Methuen, pp. xv+173, 21s.

Many readers will already have heard of Dr. Andry's recent investigation, which showed that an unsatisfactory relationship between a boy and his father is more apt to be associated with delinquency than the widely publicized factor of maternal deprivation. This volume is useful in describing his methods, his delinquent subjects and controls, his oral questionnaire, and his quantitative results, in full (a slight criticism might be raised that it is too full, since there is rather a lot of repetition). Among the most differentiating items were the boy's feeling that his mother loved him more, few activities or interests shared with the father, and failure to turn to the parents when in trouble; much absence of the father from home showed a slight association, but not the mother working. That these answers do not merely represent the delinquent's attitudes was checked by asking much the same questions of both parents in a sub-sample; there was close accord in about 80 per cent. of cases. A very consistent picture of the delinquent's conception of his father's role is revealed and Dr. Andry shows that this fits in both with Freudian, and with modern Learning, theory. At the end there is a useful short survey of the relevant literature.

CHURCHMAN, C. W., and RATOOSH, P. (editors) (1959). *Measurement: Definitions and Theories*. New York: Wiley and Sons; London: Chapman and Hall; pp. 274, 64s.

The December, 1956, meeting of the American Association for the Advancement of Science included a five-part symposium on measurement in various branches of science. The contributions to this symposium have been edited and presented as a discussion on the various views held at the moment by workers dealing with fundamental measurement. Sections in this book deal with meanings and theories of measurement and measurement in the physical and social sciences. The mathematical arguments included in the contribution by C. H. Coombs on the "Inconsistency of preferences as a measure of psychological distance," might be a useful reference to psychologists working in this very specialised section of psychometrics. In this experiment, four psychology students were presented with a scale of twelve grey colours (ranging from almost white to almost black) in sets of four at a time. It was found that each individual's concept of 'greyness' was an intermediate grey with lighter greys on one side and darker greys on the other. The necessary conditions for determining scale value for inconsistent replies is discussed.  
W. H. KING.

COLE, L. (1959). *Psychology of Adolescence*. New York: Rinehart, pp. xviii+731, \$7.00.

This is the fifth edition of a book which originally appeared in 1936. It is easy to see why this has been such a popular book: it is comprehensive, readable and helpful. There are sections on Physical, Intellectual, Emotional, Social and Moral development, and in each case, after presenting a detailed account of normal boys and girls, Dr. Cole describes the characteristics of deviates, incorporating the results of much recent research. Case studies, diagrams and statistics are widely employed and appendices suggest books, films, problems and projects.

Although the book is intended for all who have to deal with adolescents (and its publishers claim it has been "widely used as a reference by guidance counsellors, doctors, and parents, and by adolescents, themselves"), it is clear that the writer is thinking primarily of teachers, and it is teachers and students in training who would find her work particularly stimulating. Throughout the book, she brings out admirably the relevance of psychological considerations to the classroom situation. Part Six, "Educational Applications" deals with topics of perennial importance, such as "Emotional Problems in the Schoolroom," "The Social Life of the School," "The High School Curriculum."

Possibly Dr. Cole is most at ease with brilliant children. She does not distinguish sufficiently clearly between dull, backward, retarded and maladjusted—in fact, she uses 'retarded' in its more popular sense and ignores the children whose educational attainment is not in keeping with their intellectual ability whether their I.Qs. are low, average or high.

An English reader might be tempted to suggest that "Psychology of American Adolescence" would make a more accurate title. Throughout the book, Dr. Cole generalises on evidence, research and reading drawn almost entirely from American sources. In the preface, she points out that the literature on adolescence is very extensive. It is a pity that among so many hundreds of references, she did not include at least a few from outside the States. In studying adolescence (or for that matter, many other topics in psychology) the effect of a particular culture or sub-group in a culture should surely receive adequate recognition. However, an English teacher, like her American counterpart, could learn much from this book.

M. C. WRAGG.

CROTHERS, B., and PAINE, R. S. (1959). *The Natural History of Cerebral Palsy*. London: O.U.P., pp.xi+299, 32s. 6d.

In its well-balanced approach, this book makes a valuable contribution to a controversial subject. It is based on some twenty years' work with cerebral palsied children, which the authors have reviewed critically in terms of its ultimate value to the patients. In the early chapters, they discuss some of the many problems which have interested them, describe methods they have used, and suggest others which they now regret they omitted to try. They refer to relevant literature in the light of their own very considerable experience, and, in the later chapters, develop this discussion and briefly summarize their conclusions.

Their patients, admitted to the neurological wards of the Children's Hospital, Boston, between 1930 and 1950, all had motor disturbances due to non-progressive brain abnormalities derived from lesions which had occurred at an early age.

The only certain thing to emerge from a re-examination of 655 patients who were still living (17 per cent. of whom were in custodial institutions) was the necessity of studying the changing effects of the early lesions on the children during all the stages of their subsequent development. The authors are convinced that the motor disorders of these children present one only (and that often the least) of all their difficulties and that, in the frustrations and disappointments which they meet in struggling to achieve normal independence lies the main problem to be tackled.

The authors are sceptical about the value of much physical therapy in the case of the extra-pyramidal palsies of whom, in particular, an appreciable number fail to make adjustment in a competitive world primarily because of emotional and environmental difficulties. The degree to which any salvaging of these patients can



take place will, in the author's opinion, "depend on the optimism of the observer as well as on the possibility of developing means of treatment which will be more effective than those at present in use."

In a short chapter on education, the authors admit that they have "no significant information about this aspect of the situation and make no report." Nevertheless, it is obvious that they have given a good deal of thought to many of the difficulties which commonly beset those who are responsible for establishing and implementing special schemes.

Throughout the book runs a thread of warning that, since the parent-child relationship inevitably changes as each grows older, in special educational treatment as in family life, the devising of appropriate degrees of pressure and concession is one of the most difficult tasks in dealing with children whose development deviates from normal.

In conclusion, the authors state their belief that, to persist in the same sort of special treatment, whether educational or physical, which may have been effective during stages of rapid development, may be very inadvisable in later years.

The book is not entirely discursive: of the eighteen chapters, eight are devoted to classification, aetiology, pathology and treatment. The text is well illustrated with tables, figures and photographs, and a useful bibliography lists over 170 other authors to whose work reference is made.

M. I. DUNSDON.

DIENES, Z. P. (1959). *Concept Formation and Personality*. Leicester University Press, pp. xi+82, 18s.

This book describes the basis and conduct of an investigation into likely relationships between some aspects of concept formation in the mathematical field and certain personality characteristics as represented by teachers' ratings of their pupils. The early discussion offers an interesting analysis of the ability to engage in mathematical activity, distinguishing, on the one hand, the acquisition of concepts, and, on the other, the processes whereby these are manipulated to generate new and more complex ideas. It is suggested that individuals acquire and develop concepts in their own ways and not according to a stereotyped pattern. The actual processes of concept formation, as used in the later experiments, are developed from the work of Lewin, Eysenck and Rokeach. The latter part of the book describes and analyses an experiment in concept formation, along Piagetian lines with a mixed group of one hundred ten-year-old children.

Children have to learn Mathematics as it is taught in school. Moreover, concepts of number are basic to later more abstract work. After so interesting a preliminary discussion, it is, therefore, disappointing to find an experiment in which the normal school Mathematics has no place. Such an experiment is removed from both the physical and the subject context of the teaching situation, and hence two of the most important factors in the learning situation. In addition, the acquisition of such operations as used here does not correspond to what is expected of children in straightforward mathematical work. It is not known whether the ability to acquire and use these operations in this type of experimental situation would transfer into the corresponding situation involving mathematical elements. Unlike Piaget's work, this experiment seems to be concerned with the acquisition and use of a certain type of mathematical operation in a non-mathematical situation. This has probably no more relation to the learning of Mathematics than to any other subject field.

The ideas expressed in the theory are widely realised amongst teachers of Mathematics, despite the author's comments in this respect. The book is valuable in that it formalises these and brings them into perspective. But the language of the book is too complex for many teachers.

D. M. LEE.

GALANTER, E. (edit.) (1959). *Automatic Teaching—The State of the Art*. New York: John Wiley; London: Chapman and Hall; pp. viii+198, 26s.

Earlier automatic teachers, such as books, have suffered from their inability to answer back, but the latest models do so quite easily; for if the reader answers

'A' to a problem, he must turn to page 'X,' and if he answers 'B' to page 'Y,' and so on, the process being repeated until the dullest person has got all the answers right, and the brightest has pressed on through volume after volume. Alternatively, we can use 'chemo cards,' on which the right answers change colour, but the wrong answers do not, or we can construct elaborate machines. As Jacob Beck, one of the authors of this book says: "Traditional teaching devices such as books and lectures are almost completely uncontrolled. The discriminative stimuli are not explicitly specified by these devices . . . Many a difficulty in teaching can be traced to students 'focusing' on the wrong stimuli and attaching responses to them. Furthermore, texts and lectures do not specify the particular response which it is desired that an individual acquire. Since they do not generally require the student to make any response, little if anything can be known of the student's responses to the content presented . . . The rules which guide a student's responses are also indefinite . . . Reinforcement under these conditions is also a haphazard matter—depending upon accidental self-administration."

Apart from the help that such machines can obviously give the self-taught, most teachers will view mechanical aids with suspicion. Nevertheless, there is no doubt that it would benefit most teachers to have to program some of their lessons on these machines, for it becomes necessary to list the sequence of operations that an ideal teacher must perform, and to develop, in the process, a general theory of learning and teaching. We are faced, for example, with such questions as, "What is the correct order of presentation of material? . . . Is there an optimum number of errors that should be made? . . . Should prompts be used, and, if so, how many, and what kind? There is the pacing problem—experimenter-controlled rate *vs.* self-pacing; the multiple track problem . . . whether one program is satisfactory for all students and the constructive *vs.* multiple choice response problem. Many of the basic principles of learning theory have to be re-examined. For instance, is it reasonable to suppose that an aperiodic schedule may enhance the inherent interest in the subject matter, as it seems to do in animal experiments?"

This book, with its sixteen articles, twenty authors and technical vocabulary is not intended for the lay reader, but it will be understood easily enough by the professional psychologist. Chapter II "A review of factors in learning efficiency," by Gagne and Bowles, is particularly worth reading, as in addition to dealing with such topics as stimulus and response, serial tasks, interference, meaningfulness and massed *vs.* distributed learning, it also tackles the important problems of level of aspiration, task set and intrinsic *vs.* extrinsic motivation reinforcement, with separate reference to the performances involved in motor skills, rote learning, concept attainment, perceptual skills, the attainment of techniques and problem solving.

Teaching machines may, of course, prove very useful for private study, and they have the unhuman quality of infinite patience. Nor do they embarrass the shy beginner; a tape recorder, for example, may encourage the student learning to speak a foreign language to let his accent rip and to check the results better than in a class of students. There may be a danger, although it seems slight at the moment, that learning will eventually become something that one attempts "in a booth, alone," but it is surely optimistic of Howard Kendler to expect and even hope "that some dignified, elderly humanist will voice fears concerning the threat teaching machines pose, not only to the dignity of teachers, but also to the cultural standards of our entire society." They're not that good. Even with built-in Skinnerian pellet chutes they will not replace the schoolmaster.

F. W. WARBURTON.

GARRISON, KARL C., and FORCE, DEWEY, G. (1959). *The Psychology of Exceptional Children*. (Third edition.) New York: Ronald Press, pp. vi+586, \$6.00.

MAGNIFICO, L. X. (1959). *Education for the Exceptional Child*. New York, London, Toronto: Longmans Green, pp. ix+371, 35s.

The first of these books is a considerably revised edition of Garrison's textbook which was last revised in 1950. The references made to research results have been

increased and brought up to date and some topics which were only sketchily treated in the previous edition have been accorded fuller treatment, for example, cerebral palsy, certain clinical types of mental deficiency and heart conditions. The material has been re-arranged into four main sections: mental deviates (including gifted children); physical disabilities (defective vision, hearing and speech and orthopaedic defects); neurological impairments and health problems (brain damage, epileptics, delicate children); emotional and social problems (maladjustment and delinquency).

The book gathers together a good deal of basic information about the nature and causes of handicaps, summarises the psychological characteristics of exceptional children and relates these to the broad lines of education and treatment. It would, therefore, have value as a source of reference in college libraries and as an outline of the whole field for experienced teachers taking courses of training in the education of handicapped children. (It was written for this purpose in American courses in special education.) Students specialising in a particular handicap would find little in the discussion of their own field which they would not get more fully and adequately elsewhere. Moreover, the value of the book for British students is considerably reduced by the almost complete lack of reference to British work and reports, and there are the inevitable differences arising from different terminology and methods of organising special education.

These two objections apply even more to the book by Magnifico which, being largely a discussion of policy and organisation of special education in the American setting, is chiefly of interest in showing how similar problems are being discussed and tackled in another country. We should, of course, be familiar with and ready to learn from systems which differ from our own but, apart from that, this book has limited value for the British reader. Two chapters on gifted children will be of general interest. The failure of American education to give adequate attention to developing the most gifted children is discussed in relation to the philosophies and social attitudes which have impeded progress in this field. The American public's reaction to the first earth satellite is seen as an opportunity to overcome some of these barriers.

R. GULLIFORD.

GLOVER, E. (1959). *The Roots of Crime*. London: Imago, pp. xiii+422, 45s.

Dr. Glover, an outstanding authority on Freud and criminology, presents us with a lucid textbook which will rank high in its chosen field. It consists mainly of reprinted essays and lectures dating from 1922 to the present time. With faith and consistency, he has spread his concepts, at first almost alone, and from the nineteen-thirties, together with a handful of equally eminent colleagues (for instance the late Dr. Dennis Carrol) through the Portman Clinic, the Institute for the Study and Treatment of Delinquency, and since 1950, through the *British Journal of Delinquency*, of which he is co-editor with Dr. Hermann Mannheim and Dr. E. Miller.

His strong message can perhaps be best evaluated under two headings: the psycho-dynamics of criminal behaviour, and how social action can best be secured in the interest of both the criminal and of society. Both views inevitably arouse the kind of controversy to be expected when a leader pronounces on the social sciences which, as yet, are full of mysteries in comparison with already existing knowledge in the field of the normative sciences. Ultimately, the controversy is based partly on whether the pronouncements are merely hypotheses or are likely to develop into hard facts.

He stands on controversial ground, as he admits, when he states categorically that he has "no great faith in the fertility of existing forms of 'combined' scientific effort, laudable as that approach may seem to be. For... the strength of a research team is the strength of its weakest link. So long as the existence and power of unconscious motives is disregarded, we cannot learn any more than an apparent common sense dictates." Thus, he questions the efficacy of the statistical, questionnaire and prediction research approaches, which are not based on the theory of psycho-analysis.

He calls most earnestly, however, for interdisciplinary research, which embraces the very core of his and Freud's belief in the two-sided coin of the aetiology of crime

and of society's reaction to it: the existence of the super-ego as the seat of the 'unconscious conscience,' its guilt, and its unconscious need for punishment in an endeavour to relieve guilt. Through committing offences, the criminal unconsciously seeks to be punished, and, in turn, not only indirectly punishes himself, but also society. Society cries out for punishment of the criminal, because its members do not understand that their desire for vengeance stems from their own unconscious guilt feelings, which a convenient criminal scapegoat can diminish.

Given this insightful hypothesis, what is the cure? Dr. Glover makes many interesting comments, ranging from the suggestion to psycho-analyse individuals, especially the recalcitrant but not incurable psychopath, along orthodox Freudian (and not Kleinian) lines, to setting up, through the Home Office, a project for a screening service in schools, in order to detect, register and treat, all cases of pre-delinquency. Dr. Glover's important book contains many fertile ideas which should keep a generation of serious students in this field busy proving him wrong or right.

R. G. ANDRY.

HALMOS, P. (edit.) (1959). *Papers on the Teaching of Personality Development. The Sociological Review, Monograph No. 2.* Keele: University College of North Staffordshire, pp. 149. Price, 12s.

Following the issue of the first monograph of *The Sociological Review*, in 1958, which dealt with the teaching of personality development to students of education and social work, this, the second monograph, deals with further problems on Personality Development at the University of Leicester in April, 1959. In the introduction, Professor W. J. H. Sprott says that the papers of the first monograph, representing the 1958 conference at Keele, were concerned with theoretical matters, whereas the papers of this, the second conference, are severely practical in character.

The list of contributors to this volume, with the topics dealt with, is as follows: J. H. Nicholson, who writes about students' attitudes, past and present; Clare Winnicott, who deals with the development of insight; G. P. Meredith, whose subject is insight into human relations; J. W. Fibble, who presents a study of problems in the training of teachers and social workers; J. E. Richardson, who describes the exploration of teacher and pupil roles in a training house; H. F. Powell whose problem is the study of personality development; Kay F. McDougall, who presents a paper on teacher methods and problems of identification; Mia L. Kellmer-Pringle, who studies the teaching of personality development in a diploma course for experienced teachers; Norah Gibbs, who reflects on the conferences at Keele and Leicester; and the contributions are brought to a close by Paul Halmos, who discusses the technical problems in teaching dynamic psychology and the moral issues of professional education.

It is invidious to single out any of these papers for special mention, but those by Clare Winnicott, Norah Gibbs, and Paul Halmos appear to the reviewer to deserve careful re-reading. The others are all undoubtedly of very much interest, but in different ways and to different kinds of readers, so that the whole volume is one deserving the attention of any persons concerned with the teaching of personality development and its problems to students of education and the social sciences. The editor is to be commended on his success in collecting such an interesting group of contributors.

R. W. PICKFORD.

KELLMER-PRINGLE, M. L., and SUTCLIFFE, B. (1960). *Remedial Education—An Experiment.* Birmingham Printers, pp. 36, no price given.

Maladjusted children in the Caldecott Community tend to be seriously backward educationally, though not necessarily below average in Terman-Merrill I.Q. The junior author was encouraged to set up a remedial clinic in the Community where, for two years, an hour's help a week with reading or arithmetic (or both) was given individually to some twenty children, plus a small number of local authority children in care. Complete details of the test scores are given and these appear to show



excellent progress, though one feels some doubts about the method of calculating 'acceleration.' For example, a child who had reached a Reading Age of 6.0 after five years at an ordinary school and 8.0 after two years' attendance at the clinic, is claimed to have increased his rate of progress five times.

A novel feature of the Report is that good progress in reading (though not in arithmetic) was maintained for several months after discharge. The pamphlet also contains a useful discussion of the educational difficulties of the emotionally disturbed child, and of the place of such a remedial unit in schools or homes for the maladjusted.

KNIGHT, R., and KNIGHT, M. (1960). *A Modern Introduction to Psychology*. (Sixth edition, revised). University Tutorial Press, pp. 272, 9s.

The Sixth Edition, Revised, of this extremely concise and admirably lucid introduction to psychology is to be welcomed. The first five chapters have been thoroughly revised as have the chapters on "Animal Learning" and "Instinct." The chapter on "Brain and Mind" has been re-written to give more adequate insight into the semantics of the topic. The Notes on Reading have been brought up to date and some new footnotes have been added.

The balance of the book remains weighted in favour of physiological psychology, which takes up about half the text. The remainder is given over to Intelligence, Learning, Personality, Instinct and Psycho-analytic Theory. It might well have been more thorough to have included a chapter on "Social Psychology." This could have replaced the chapter on "Dreams" which seems the least satisfactory chapter.

Good though this introductory book is, it may be more suitable for students of psychology than for students of educational psychology. Nevertheless, it should be included in their recommended reading because of the use the authors make of relevant research and the clarity with which they write.

P. H. TAYLOR.

LEANING, P. A. (1959). *The Challenge of Cerebral Palsy*. Auckland, N.Z.: Whitcombe and Tombs, pp. 98, no price given.

The author intends this book for teachers and parents of cerebral palsied children. For those coming newly to the field, it will provide a useful introduction. Miss Leaning gives a concise account of the physical and educational handicaps of cerebral palsied children, and discusses the remedial methods employed. The main part of the book is devoted to educational problems and methods, and there are also chapters on speech, occupational and physio-therapy, as well as vocational guidance and training. Before writing this book, Miss Leaning made a world-wide tour of cerebral palsy centres, and so she is able to illustrate her points with brief accounts of some of the different educational approaches she observed, and with the comments and opinions of those she met.

Those who are familiar with this field will probably gain most from Miss Leaning's account of the effect of the physical handicap on the child's personality. She is herself cerebral palsied, and so is able to bring to her very objective treatment of this subject, the insight gained from her own experience. She points out how the effect of the physical handicap extends to many unsuspected facets of the child's life, and emphasizes the importance of parents' and teachers' attitudes in avoiding a feeling of hopelessness in the child.

Readers of this book already working for cerebral palsied children will welcome Miss Leaning's plea for a greater exchange of information in this field, and will only regret that she does not give more detailed information about the centres she visited. Perhaps, however, this will follow.

K. WEDELL.

MADDOX, H. (1960). *Maddox Verbal Reasoning Test*. Edinburgh: Oliver and Boyd, Tests 7d. each, Manual 1s. 9d.

Almost all group intelligence tests for children below 10½ use the Mental Age method of scoring, thus this new one will be a welcome addition for testing third-

year juniors. It closely resembles a Moray House Verbal Reasoning test with deviation-quotient norms, though in fact, it does not register below I.Q. 78 at 9½ years. It was standardised on some 4,500 children in the Midlands. The one-hundred items are of conventional type, and 45 minutes are allowed. It is preceded by a practice sheet taking about 15 minutes, but thereafter, the children have to read the instructions and items for themselves. The author admits its dependence on reading, and prefers to regard it as a measure of 'scholastic aptitude' rather than 'intelligence.'

MOHSIN, S. M. (1959). *Practical Handbook of Guidance in Secondary Schools*. Bihar, Patna: Superintendent Secretariat Press, pp. 169, Rs. 2.25.

The Bihar State Bureau of Educational and Vocational Guidance, under the directorship of the author, appears to have set up an interesting cross between the American system of school counselling and the British system of youth employment service. In each secondary school a teacher is given 6 months' training in the theory and practice of counselling. He then becomes responsible for selecting 13-7 pupils for transfer from elementary schools. He keeps records and applies attainments tests during their first year as a basis for allocating them to one of the various secondary courses—science, arts, technical, agricultural, etc. Throughout the school life, he is available for consultation on scholastic and personal problems, and is responsible for providing occupational information at the appropriate time. Then, before leaving age at 17+, he has complete records of the school career and background, and can apply interests and aptitude tests (mostly adapted from American sources by the Bureau). Finally, he refers the leaver to the Youth Employment Office of the Employment Exchange, with his recommendations.

It is not made clear how completely this system operates, but it would certainly appear to be a sound and useful scheme—better than what we have here with our (largely untrained) careers masters. This volume is a working textbook for counsellors, covering the main points in the organization of their work and principles of counselling.

NICHOLSON, R. J., and GALAMBOS, P. (1960). *Performance in G.C.E. (Advanced Level) Examinations and University Examinations*. University of Hull, Occasional papers, pp. 22, 2s.

This investigation is similar to several other follow-up studies of the predictive value of secondary school examinations for university courses, and yields the same depressing results. Among groups of around a hundred students in the University of Hull, the correlation between A-level and first-year results in English was .16; Economics .22; and with Finals the figures were even lower. However, in French and Chemistry the coefficients reached .33 and .41. But these, as the authors point out, hardly justify the use of A-level passes as the main basis for admission.

There are several other points of interest to those concerned with university selection. The reviewer would disagree with the author's rejection of corrections for homogeneity; for surely what one wants to know is the success of A-level in differentiating *candidates* who will perform reasonably at university from those who are rejected because they are likely to get very low marks, rather than in differentiating the better from the poorer *acceptances*.

OPIE, I., and OPIE, P. (1959). *The Lore and Language of Schoolchildren*. Oxford University Press, pp. xix+417, 35s.

This is a fascinating record of patient research on 5,000 children in different parts of England, Scotland and Wales, attending seventy schools of all types, and improprieties, nicknames and superstitions, jeers against teachers, and procedures of ordeals through which children are put. The close resemblance of many of these items in such far separated parts of the country is remarkable, though local influences are often discernible. In many cases origins can be traced back to the middle ages

or earlier; in others the children have themselves adapted an old expression or rhyme to fit a new situation. Even links with the U.S.A. are revealed.

The book is more likely to be useful to the sociologist than to the psychologist, though the latter may profit from noting the prevalence of compulsive-obsessions in presumably normal children. I was glad to add to my own collection of these the reference to General Gordon's habit of zig-zagging on a stone pavement to avoid treading on the cracks.

Certainly, this volume would make at least an excellent bed-side book for the child psychologist.

C. W. VALENTINE.

PETERS, R. S. (1959). *Authority, Responsibility and Education*. Allen and Unwin, pp. 137, 12s. 6d.

"One of the most awe-inspiring features of academic institutions is the tyranny of fashions," as Dr. Peters wryly observes. Any book that leads its readers to examine the roots of current educational beliefs and practices is, therefore, salutary, whether the reader accepts the author's findings or not.

Dr. Peter's book, based mainly on radio talks given between 1956 and 1959, does just this. Some of the educational and philosophical concepts he scrutinises are age-old, others are of more recent appearance, popular derivations of Marx and Freud, but all concern fundamental educational issues.

This radical approach leads him not only to question common assumptions, but sometimes to make a case for their opposite. He suggests, for instance, that far from being cleverer than he is moral, twentieth-century man has so far not shown enough intelligent foresight to ensure the maintenance of old morality in new social conditions. He also suggests that the psychology of learning could benefit more from a study of the practice of education than the teacher can at present from learning theory.

But Dr. Peters builds most of his case for the latter on the mechanistic learning theories of Hull and Skinner, and omits more recent work on learning in its genetic and social setting (except that of Piaget) which in the experience of this reviewer some teachers find both relevant and useful. Here Dr. Peters falls into the same sort of error as he has in previous parts of the book been spotlighting and castigating in others.

Nevertheless, the book makes lively reading, and it is recommended to lecturers in Education, not as a complete statement of the case on any of the points raised, but as a stimulating point of departure for seminar discussion groups.

CORA TENEN.

PHILLIPS, B. N., DUKE, L. D., and DE VAULT, M. V. (1960). *Psychology at Work in the Elementary School Classroom*. New York: Harper, pp. xii+395. No price given.

Both in its method and its content this is an unusual approach to the teaching of educational psychology. Essentially, it consists of a series of reprints of recent research articles dealing with aspects of child behaviour in classroom situations. Each article is preceded by a brief survey of the wider field which is represented by the particular research together with a selected list of references and a discussion of the underlying psychological principles. The application of these principles to educational problems is emphasised. The papers are well-chosen 'run of the mill' enquiries carried out mainly in America during the past ten years. They make use of a wide range of experimental methods and statistical treatments, and are classified on their content under two main headings, "Classroom Organisation and Interpersonal Relationships" and "Pupil Behaviour and Learning."

On the assumption that transfer from research findings to teaching is not automatic, each group of original articles is followed by a short section demonstrating the relationship between experiment and practice. These sections consist of a number of fictitious classroom situations which are interpreted in the light of the principles derived from research so that points of transfer are made abundantly clear.

There is no doubt a need to demonstrate to teachers the relevance for the classroom of some of the results of research and attempts to do so are to be welcomed. It is difficult, however, to see how this book will succeed with teachers and students in this country. This is mainly because so few of them have sufficient training in psychology and statistics to make much of the research papers. It is true that many of the papers have a "Summary and conclusions" paragraph at the end but this, by itself, would have little value. There seems, in fact, to be only one way of making research findings useful to teachers, and that is to teach them much more psychology and statistics during their initial and in-service training, so that they can at least read with understanding the publications which have a bearing on their craft.

L. B. BIRCH.

PRESSEY, S. L., ROBINSON, F. P., and HORROCKS, J. E. (1959). *Psychology in Education*. New York: Harper, pp. xi+658, \$6.50.

MACDONALD, F. J. (1960). *Educational Psychology*. San Francisco: Wadsworth Publ. Co., pp. xix+748, no price given.

The first of these books, written by three distinguished members of Ohio State University, brings up to date, without greatly extending, a work written by Pressey in 1933. It is of broad coverage without being overloaded with detail and includes relevant research up to 1957.

It departs from the conventional approach in treating development as a process not wholly concerned with childhood and so does not require separate chapters on adolescence. In its second part there is a well organized treatment of the nature and fostering of learning, and in the third, a somewhat extended treatment of working with the individual student. One chapter of this, "Bettering the Best," reflects the change of emphasis now becoming apparent in certain areas of the United States. As a well written, well produced, discursive text which keeps within the chosen field, it should be within the reading level of first and second year training college students and should certainly find a place within college libraries.

Macdonald's textbook, on the other hand, is not so easy to read, varies in its quality, is repetitive in parts, and uses a variety of print sizes and diagrams which are so schematic and heavily printed as to make them appear odd. Nevertheless, it does two things which few texts do—it presents a point of view about educational psychology, and it challenges the reader. The text itself elaborates views expressed by Colardarci at both Indiana and Stanford, and devotes ten out of its eighteen chapters to "Learning and the Development of Personality."

Its viewpoint finds clearest expression in the third chapter, "The Learner and the Educative Process," and the final one—"The Teacher and the Improvement of Educational Practice." Every action which the teacher takes, every method or procedure used, is based on a hypothesis about behaviour change. Macdonald aims "to make teachers become more intelligent hypothesizers and to encourage them to analyze critically the bases for the decisions they make." Throughout he stresses that teachers must "do something." Each action, or method, is based upon some interpretation of knowledge, whether the interpretation is explicit or not. Some teachers lack information about research conducted by psychologists: he attempts to remedy this, not by giving a very great number of references but by describing a more limited number of researches in sufficient detail for the reader to draw his own conclusions and to attempt his own deductions and generalizations. Other teachers may base their methods on imperfectly understood pieces of research: these, he attempts to help by discussing other implications of the research quoted, and by an examination of the conditions under which generalizations may be successful and under which they may fail.

Now that the three-year training college course is nearly with us, and that students are to become more mature before they emerge as teachers, this becomes an excellent book to use with these students in their third year. For them, and their own teachers, it will more than repay close and continuing study, and should do much to rescue postulant teachers from mere subscription to rule of thumb methods employed without regard to context or situation.

R. EDWARDS.



STOGDILL, R. M. (1959). *Individual Behaviour and Group Achievement: A Theory—The Experimental Evidence*. New York: Oxford University Press, pp. 352, 40s.

The author has already published a large number of researches on sociometry, leadership and group relationships during the past ten years and this book reports a theory which he has developed in the course of his close association with the topic of social groups. He describes an organised group as an input-output system in unstable balance, and proceeds to analyse the structure of the 'input' and 'output' variables. He assumes that the minimum aspects of behaviour necessary to describe 'input' are the performances, interactions and expectations of the members of the group. The last, which would not necessarily be allowed by other theorists, is a variable which is associated with the measurement of the purpose of the group; formally he defines 'expectation' as "readiness for reinforcement, and is a function of drive, estimated probability of a possible outcome and the estimated desirability of this outcome." He comments on the obvious links between this definition and modern learning theory. He also criticises a tendency of some writers to confuse 'performance'—an individual variable, with 'interaction'—a group variable. These three variables, in combination, "account for the development of group structure and for the initiation and maintenance of group operations." These, in turn, result in 'output' or achievement, in which he includes not only productivity, but morale and integration. He has an interesting concept of group morale as "the degree of freedom from restraint exhibited by a group in operating on a goal objective." These three dimensions of output are not necessarily independent.

Within this framework, a substantial part of the book is devoted to a detailed examination of each of the input and output variables and their relationships both to each other and to the internal working and development of the group. In each case the theory is followed by an examination of available research findings for confirmatory evidence. The lack of evidence from experiments scientifically designed to test the various aspects of the theory is a weakness which the author himself realises. He states, in fact, "... some may doubt that the same experiment can lend equal support to various theories and sub-theories." The connection between the research reported and theory is often tenuous and the thread in the author's mind often tends to be obscured by the enormous amount of detail he includes.

In discussing 'interaction' he attempts to find ground common to his theory of social groups and the mathematical theory of groups. This is not explicitly developed, and its introduction perhaps seems a little pretentious. Nevertheless, for the advanced student of group theory the book is likely to be of use on account of its wide coverage of existing research in the form of abstracts, as well as the collected bibliography. It would have made its points more cogently if the author had made some effort to vary the method of presentation with the aid of tabulations and charts.

D. M. EDWARDS PENFOLD.

TANSLEY, A. E., and GULLIFORD, R. (1960). *The Education of Slow Learning Children*. Routledge and Kegan Paul, pp. viii+255, 28s.

This book will be invaluable to all teachers in E.S.N. schools, and it has much to offer to those concerned with backward classes or groups in ordinary schools. A large amount of practical guidance is given on methods of helping the very backward in language and reading, number, arts and crafts, and general social competence; and many ingenious ways are described of enlisting their interest and building up favourable attitudes after a long history of failure and frustration.

The first four chapters will chiefly interest psychologists and students of education. The first surveys the present facilities for education of the backward, and brings out the unsatisfactoriness of using E.S.N. schools as a dump for pupils ranging from retarded unstabiles to ineducables. It presses the needs both for better facilities for backward groups in ordinary schools, and for flexibility in transferring children from one type of school or class to another. The second chapter puts forward a

sensible position on intellectual deficiency, bringing out the value (though also the limitations) of intelligence measures, while admitting that I.Q.s are often lowered by unfavourable environment and may alter, and avoiding the fallacies of the Achievement Quotient. The third shows how the needs of E.S.N. children differ but little from those of the average child, though they are frequently characterised by emotional immaturity. The fourth is a useful compendium of information on the common physical deficiencies often associated with intellectual backwardness.

Apart from the chapters on teaching, perhaps the most original section of the book is the last one which discusses ways of preparing the E.S.N. leaver for a vocation and for life as a self-supporting adult. P.E.V.

TIBBLE, J. W. (1960). *Conference Report*. London: New Education Fellowship, pp. 84, 5s. 6d.

This is a brief report on the conduct of an international Conference held at Utrecht in 1956, where most of the work was done in discussion groups under experienced group leaders. A questionnaire was circulated eighteen months later, and a majority of the members who answered seem to have had favourable recollections. The useful effects of the procedure and occasional weaknesses are carefully analysed and discussed. Curiously, the report seems to have nothing to say as to whether members learnt anything fresh about the ostensible topic of the Conference—Constructive Education and Mental Health in Home, School and Community.

WEAVER, A. (1959). *They Steal for Love*. (London: Max Parrish, pp. 132, 12s. 6d.)

This book describes an experiment in Education and Psychiatry with children and parents. Facilities were made available under which a small number of disturbed children could be brought together to live in a community which was nearly as possible a normal one in the generally accepted sense of that term. The community was headed by the author, who acted as Warden, together with a small staff of housemasters, a teacher and a visiting psychiatrist. The original intention was to help pre-delinquent and disturbed children on a short-stay basis. Largely for administrative reasons, this did not prove possible, and the children tended to be drawn from more severe and long-standing cases. An account is given of the different aspects of the community and their management. Case-histories of both parents and children are discussed. Some follow-up studies of the children are also presented.

The most important feature of this work is the extent to which it succeeds in bridging the gap which too often exists between general psychological or psychiatric theory and educational or social practice. From the educational angle, the author presents an informal and concise account of the course of the experiment. In human terms, there is the careful study of a number of individuals, showing the kind of light is thrown on a number of corresponding psychological concepts discussed by such writers as Suttie, Bowlby, Flugel and Mead. The psychological and psychiatric relevant general concepts as their description proceeds. The importance and depth of these concepts in the processes of day-to-day living is demonstrated in both family and community life.

This kind of analysis is of very great interest to parents and teachers, as well as to those who work with children and adolescents in more specialised fields. It is essential that children such as those described should be helped with their difficulties as early as possible. Parents and teachers should try to realize how fine is the division between these children and their counterparts in the normal classroom, and yet to give in these respects. It should be read by all students in training as well as by qualified teachers and child workers. It will enrich their knowledge of child development in both the classroom and the wider world.

D. M. LEE.

WOOTTON, BARBARA (1959). *Social Science and Social Pathology*. London: George Allen and Unwin, pp. 400, 35s.

This book offers us a Social Scientist's analysis of present-day views and attitudes on anti-social behaviour and all that they imply. The first part of the book contains a useful review of research on the relation between criminality and various social and economic factors, e.g., truancy, broken home, poverty, etc., and critically examines the validity of certain currently popular hypotheses.

One of these is the theory of maternal separation as a causative factor in crime. Lady Wootton, after reviewing the contradictory evidence, concludes that all that has been shown is that children need dependable love, something which she, not unnaturally, regards as a homely truth known for centuries. The main usefulness of the theory, she claims, lies in throwing light on "deplorable patterns of institutional upbringing." Is this not both an over-simplification of the theoretical position as regards the fundamental early relationships and an underestimate of the practical consequences the wide acceptance of the theory has had—which would not have come about without?

The second part deals with changing contemporary attitudes towards social deviants, so deeply coloured by the psychiatric approach. While Lady Wootton welcomes the humanising influence of this trend, this part—and indeed, the book—is largely devoted to a critical exploration of the limitations and fallacies of psychiatric theories. (Psychologists are presumably included among the "psychiatrically-orientated satellites.")

The book renders a service in drawing attention to the comparatively subjective nature of the definitions of mental health and mental illness (the author has amassed an impressive list of them), yet it also seems to exaggerate their inadequacy. Lady Wootton assumes that mental illness is generally diagnosed merely on the basis of delinquent behaviour. When reviewing present practice, she writes: "It is almost true to say that the illness is the behaviour for which it is also the excuse" (page 225). She here ignores the possibility that an unresolved conflict within the self may be regarded as *being* the illness and that this is diagnosed by a syndrome of particular dys-functions, not all of which are of a delinquent nature.

In an excellent chapter on "Mental Disorder and Criminal Responsibility," she explores with relentless logic the difficulties involved in arriving at clear criteria for diminished responsibility in law other than the intellectual criteria of the M'Naughten Rules, while still retaining the concept of responsibility in any form. She herself seems to favour the policy of by-passing the question of responsibility and deciding the kind of 'treatment' to be meted out to the criminal by assessing what will help most to modify his attitudes.

Part III of the book contains the conclusions on methodology, where Lady Wootton favours carefully controlled research with modest aims ('prediction' rather than 'causation'), and practical conclusions which are less clear-cut. She prefers the environmental to the individual approach to social problems because she fears that the present emphasis on individualistic study will divert attention away from the need for environmental action, in fact, that psychological treatment will become the new opium of the masses. This view makes her sceptical of Child Guidance Clinics. She considers that it would be more logical to modify the régime in the normal educational system rather than to create new institutions to deal with its misfits. Dealing with misfits, one feels, far from covers all the functions of a Child Guidance Clinic, one of which surely is precisely to guide and help schools to adopt a psychologically sound approach to children.

Throughout the book there runs a note of regret at the inordinate expansion of the medical empire in these fields, and, indeed, it may well be pondered whether defects of social and interpersonal development and their treatment need of necessity be the prerogative of medically qualified persons. (Psychiatrists like Dr. Ian Suttie and Professor Aubrey Lewis seem to have queried this.)

The book radiates common sense, in lucid and most readable prose. It questions the foundations of many of our assumptions and exposes implications of our attitudes that we do not always face up to. We are indebted to Lady Wootton for making us think.

H. LYTTON.

ZANGWILL, O. L. (1960). *Cerebral Dominance and its Relation to Psychological Function*. Published for the William Ramsey Henderson Trust. Edinburgh: Oliver and Boyd, pp. vii+31, 10s. 6d.

In the earlier part of the book Professor Zangwill presents evidence suggesting that in left-handed, or predominantly left-handed individuals, the dominant hemisphere as usually ascertained is more often the left than the right. Furthermore, from the work of Luria, also that of Subirana, it seems that patients who are left-handed, or have left-handed relatives, are more likely to have complete recovery from aphasia than dextral patients. On the other hand, the evidence also suggests that language mechanisms are less fully lateralized in sinistrals than dextrals, and in the majority of left-handed individuals speech mechanisms are likely to be disturbed by an appropriate located lesion in either cerebral hemisphere.

During the latter parts of the work the author examines the findings from twenty cases of backwardness in reading and spelling in which there was no evidence of primary mental subnormality or gross brain damage. The results suggested to him that there is a *prima facie* case for linking dyslexia and kindred disorders with a sinistral tendency in the individual or in the family. Zangwill makes it clear, of course, that left-handedness and ambidextrality as such are not causal factors in the aetiology of dyslexia. But he does propose, from his own findings, and from those of other workers, that some left-handers, some who are ambidextrous, and some right-handed individuals who have sinistral antecedents or a covert sinistral tendency, may possess a type of cerebral organisation which is particularly vulnerable to the effects of stress. Thus, in these cases reading and other language functions may easily be disturbed by environmental conditions.

To the reviewer it seems that the clear opinion of such distinguished workers as Penfield and Roberts that speech mechanisms depend upon the integrity of the left hemisphere in almost all cases, regardless of handedness, gives rather more emphasis to the importance of the left hemisphere than does the evidence marshalled in Table 3. But Professor Zangwill is well aware of this from his remarks at the bottom of page 9.

This book is easy to read and should be studied by all who are concerned with backwardness in reading.

K. LOVELL.

### PUBLICATIONS RECEIVED

The mention of a book in this list neither implies, nor precludes, a later review.

- C. I. HOVLAND and I. L. JANIS (ed.) (1959). *Personality and Persuasability*. Yale Studies in Attitude and Communication, Vol. 2. New Haven: Yale University Press, pp. xiv+333, 40s.
- H. DIACK (1960). *Reading and the Psychology of Perception*. London: Peter Skinner, pp. 155+xxiii, 21s.
- B. M. BASS (1960). *Leadership, Psychology and Organizational Behavior*. New York: Harper, pp. xiii+548, \$6.50.
- H. M. ROBINSON (edit.) (1959). *Reading Instruction in Various Patterns of Grouping*. University of Chicago Press Supplementary Educational Monographs No. 89, pp. iv+212, 28s.
- C. B. HINDLEY, E. IRVINE and E. MILLER (edit.). *Journal of Child Psychology and Psychiatry and Allied Disciplines*. London: Pergamon Press, Vol. 1, No. 1, Jan. 1960; to appear quarterly. Annual subscription £5 5s. (private individuals), £7 (libraries and institutions).
- L. J. STONE and J. CHURCH (1957). *Childhood and Adolescence*. New York: Random House, pp. xvii+456, no price given.



- C. V. BURGESS (1960). *The Burgess Plays*, Books 1-4. University of London Press, pp. 32, 1s. 3d. each.
- A. E. TANSLEY and R. GULLIFORD (1960). *The Education of Slow Learning Children*. London: Routledge and Kegan Paul, pp. viii+255, 28s.
- L. MUKHERJEE (1960). *Problems of Administration of Education in India*. Allahabad Kitab Mahal, pp. ix+296, 21s.
- D. G. BURNS (1960). *Vocabulary of the Secondary Modern School Child*. London: National Foundation for Educational Research, Occasional Publication No. 1, pp. 175, 8s. 6d.
- H. I. ABELSON (1960). *Persuasion: How Opinions and Attitudes are Changed*. London: Crosby Lockwood, pp. ix+118, 27s.
- F. MCKINNEY (1960). *Psychology of Personal Adjustment* (3rd edit.). New York: John Wiley; London: Chapman and Hall, pp. xiii+490, 52s.
- B. R. BUGELSKI (1960). *An Introduction to the Principles of Psychology*. New York: Rinehart, pp. xvii+571; Instructor's Manual, pp. 27, \$6.00.
- M. L. KELLMER-PRINGLE and B. SUTCLIFFE (1960). *Remedial Education—An Experiment*. University of Birmingham, Education Department, pp. 36, 2s.
- A. L. EDWARDS (1960). *Experimental Design in Psychological Research*. (Rev. edit.). New York: Rinehart, pp. xiii+398, \$6.50.
- A. B. ROSS (1960). *The Education of Childhood*. London: Harrap, pp. 159, 10s. 6d.
- G. F. FARWELL and H. J. PETERS (1960). *Guidance Reading for Counselors*. Chicago, Ill.: Rand McNally, pp. x+691, no price given.
- V. SOUTHGATE and J. HAVENHABD (1960). *Sounds and Words*, Pupils' Books 1-6 and Teachers' Book. Prices 2s. 3d. each, and 2s. 6d.
- A. W. G. EWING (1960). *The Modern Educational Treatment of Deafness*. Manchester University Press, 45s.
- S. B. SARASON, et al. (1960). *Anxiety in Elementary School Children: A Report of Research*. New York: John Wiley, pp. viii+351, no price given.
- H. MADDOX (1960). *Maddox Verbal Reasoning Test*. Edinburgh: Oliver and Boyd, Manual 1s. 9d., tests 7d. each.
- O. L. ZANGWILL (1960). *Cerebral Dominance and its Relation to Psychological Function*. Henderson Trust Lecture No. XIX. Edinburgh: Oliver and Boyd, pp. vii+31, 10s. 6d.
- R. S. ILLINGWORTH (1960). *The Development of the Infant and Young Child, Normal and Abnormal*. Edinburgh: Livingstone, pp. viii+318, 27s. 6d.
- S. M. MOHSIN (1959). *Practical Handbook of Guidance in Secondary Schools*. Bihar, Patna: Superintendent Secretariat Press, pp. 169, Rs. 2.25.
- J. HEMMING (1960). *Problems of Adolescent Girls*. Heinemann, pp. ix+179, 18s.
- I. D. SUTTIE (1960). *The Origins of Love and Hate*. Penguin Books, pp. 222, 3s. 6d.
- D. J. WEST (1960). *Homosexuality*. Penguin Books, pp. 200, 3s. 6d.
- W. S. LLOYD (1959). *The Role of Government in Canadian Education*. Toronto: Gage, pp. 98, no price given.
- F. C. BOON (1960). *Companion to School Mathematics*. London: Longmans, pp. 301, 30s.
- C. SELLTIZ, M. JAHODA, M. DEUTSCH, S. COOK (1960). *Research Methods in Social Relations*. London: Methuen, pp. xiii+621, 42s.
- W. C. TROW (1960). *Psychology in Teaching and Learning* (Book and Manual). U.S.A.: Houghton Mifflin Company, pp. xv+488, \$5.75.
- H. GULLIKSEN and S. MESSICK (1960). *Psychological Scaling*, pp. xv+211, no price given.

# INDEX TO VOLUME XXX, 1960

Articles :	PAGE
Ability and Attainment of Children at Three Age Levels—A National Survey of: D. A. PIDGEON .. .. .	124
Aggressive and Withdrawing Children in the Normal School. I—Patterns of Behaviour; II—Disparity in Attainment: E. A. LUNZER .. .. .	1, 119
Attainment, Differences in, between Primary Schools in Mixed-Language Areas: Their Dependence on Intelligence and Linguistic Background: D. G. LEWIS .. .. .	62
Attainments of Educable Subnormal Adolescents, Levels of: R. E. MORAN .. .. .	201
Bias in Marking Examination Scripts, A Type of: M. J. FARRELL and N. GILBERT .. .. .	47
Bilingualism and Non-verbal Intelligence, A Critical Study of: W. R. JONES .. .. .	71
Class Groups of Contrasting Climate, Behaviour in: D. V. CONNOR .. .. .	244
Conservation of Substance in the Junior School Child, A Study of the: K. LOVELL and E. OGILVIE .. .. .	109
Definitions of Words, A Note on the Responses Made by Secondary School Children in their: D. G. BURNS .. .. .	30
Engineering Apprentices in Part-Time Technical College Courses, Placement Problems Among. Part I—Range of Ability: E. C. VENABLES .. .. .	237
English Language, The Performance in, at 'O' Level, of a Sample of University Students: E. A. HEWITT .. .. .	40
Exhibitions, Methods of Assessing the Value of: D. SHEPPARD .. .. .	259
Goal-setting Behaviour, Anxiety and School Streaming: T. VENESS .. .. .	23
Heredity and Environment in regard to 'Measured Intelligence,' Interaction of: D. H. STOTT .. .. .	95
Personality Assessments and Academic Performance in a Boys' Grammar School: E. ASTINGTON .. .. .	225
Psychology in Training Colleges, The Teaching of: K. G. COLLIER .. .. .	103
Research on Allocation to Secondary Education, Reflections on: R. R. DALE .. .. .	220
Retest Discrepancy in Mentally Sub-normal Children, Observations on: D. H. STOTT .. .. .	211
Selection by Procedures of Varying Validities, The Measurement of, Efficiency of: F. SANDON .. .. .	134
Spatial Ability, The Validity of Tests of, as Predictors of Success in Technical Courses: I. M. SMITH .. .. .	138
Symposium: Contributions to the Diagnosis and Remedial Treatment of Reading Difficulties:	
I—The Investigation of Reading Problems To-day: M. D. VERNON .. .. .	146
II—The Effect of Practice on Performance in Scholastic Tests: W. CURR and N. GOURLAY .. .. .	155
III—Part-time Classes for Young Backward Readers: H. H. HILLMAN and R. L. SNOWDON .. .. .	168
IV—The Effects of Remedial Education on Young Children's Reading Ability and Attitude to Reading: J. DUNHAM .. .. .	173
V—A Study of Motivation in Remedial Reading: G. R. ROBERTS .. .. .	176
Symposium: The Development of Moral Values in Children:	
VII—The Contribution of Learning Theory: H. J. EYSENCK .. .. .	11
VIII—Freud's Theory of Moral Development in Relation to that of Piaget: R. S. PETERS .. .. .	250
Zero Error in Moray House Verbal Reasoning Tests: A. E. G. PILLINER, J. SUTHER- LAND and E. G. TAYLOR .. .. .	53
Comments:	
Comments on Articles by Mr. D. G. Lewis and Mr. W. R. Jones: J. L. WILLIAMS .. .. .	271
Replies: D. G. LEWIS and W. R. JONES .. .. .	272
Comments on Dr. Bene's Article: W. B. DOCKRELL .. .. .	78
Reply: E. A. BENE .. .. .	80
On D. H. Stott's Article: Interaction of Heredity and Environment in regard to 'Measured Intelligence': C. BURT .. .. .	273
Reply: D. H. STOTT .. .. .	275
Reply to: A Critical Study of Bilingualism and Non-Verbal Intelligence (by W. R. Jones): D. G. LEWIS .. .. .	78
Critical Notice:	
The Crowther and Albemarle Reports: C. BURT .. .. .	277

# Research Note :

The Peabody Picture Vocabulary Test with English Children : J. W. Moss and P. EDMONDS .. .. .	82
---	----

# Summaries of Research Theses :

By N. W. KING, E. POTTS, P. N. P. WILLIAMS, J. B. MANNIX, P. V. REDDY, N. ROBINSON, C. C. TAYLOR .. .. .	82, 180, 266
--	--------------

# Authors :

ASTINGTON, E. : Personality Assessments and Academic Performance in a Boys' Grammar School .. .. .	225
BENE, E. A. : Reply to Comments by Professor Dockrell .. .. .	80
BURNS, D. G. : A Note on the Responses made by Secondary School Children in their Definitions of Words .. .. .	30
BURT, C. : On D. H. Stott's Article : Interaction of Heredity and Environment in Regard to ' Measured Intelligence ' .. .. .	273
— Critical Notice : The Crowther and Albemarle Reports .. .. .	277
COLLIER, K. G. : The Teaching of Psychology in Training Colleges .. .. .	103
CONNOR, D. V. : Behaviour in Class Groups of Contrasting Climate .. .. .	244
CURR, W., and GOURLAY, N. : The Effect of Practice on Performance in Scholastic Tests .. .. .	155
DALE, R. R. : Reflections on Research on Allocation to Secondary Education .. .. .	220
DOCKRELL, W. B. : Comments on Dr. Bene's Article .. .. .	78
DUNHAM, J. : The Effects of Remedial Education on Young Children's Reading Ability and Attitude to Reading .. .. .	173
EYSENCK, H. J. : Symposium : The Development of Moral Values in Children. : VII—The Contribution of Learning Theory .. .. .	11
FARRELL, M. J. and GILBERT, N. : A Type of Bias in Marking Examination Scripts .. .. .	47
HEWITT, E. A. : The Performance in English Language, at ' O ' Level, of a Sample of University Students .. .. .	40
HILLMAN, H. H., and SNOWDON, R. L. : Part-time Classes for Young Backward Readers .. .. .	168
JONES, W. R. : A Critical Study of Bilingualism and Non-Verbal Intelligence .. .. .	71
— Reply to Comments by J. L. Williams .. .. .	272
LEWIS, D. G. : Differences in Attainment between Primary Schools in Mixed-Language Areas : Their Dependence on Intelligence and Linguistic Background .. .. .	63
— Reply to : A Critical Study of Bilingualism and Non-Verbal Intelligence (by W. R. Jones) .. .. .	78
— Reply to Comments by J. L. Williams .. .. .	272
LOVELL, K., and OGILVIE, E. : A Study of the Conservation of Substance in the Junior School Child .. .. .	109
LUNZER, E. A. : Aggressive and Withdrawing Children in the Normal School. : I—Patterns of Behaviour .. .. .	1
II—Disparity in Attainment .. .. .	119
MORAN, R. E. : Levels of Attainment of Educable Subnormal Adolescents .. .. .	201
MOSS, J. W. and EDMONDS, P. : The Peabody Picture Vocabulary Test with English Children .. .. .	82
PETERS, R. S. : Freud's Theory of Moral Development in Relation to that of Piaget .. .. .	250
PIDGEON, D. A. : A National Survey of the Ability and Attainment of Children at Three Age Levels .. .. .	124
PILLINER, A. E. G., SUTHERLAND, J., and TAYLOR, E. G. : Zero Error in Moray House Verbal Reasoning Tests .. .. .	53
ROBERTS, G. R. : A Study of Motivation in Remedial Reading .. .. .	176
SANDON, F. : The Measurement of Efficiency of Selection by Procedures of Varying Validities .. .. .	134
SHEPPARD, D. : Methods for Assessing the Value of Exhibitions .. .. .	259
SMITH, I. M. : The Validity of Tests of Spatial Ability as Predictors of Success in Technical Courses .. .. .	138
STOTT, D. H. : Interaction of Heredity and Environment in regard to ' Measured Intelligence ' .. .. .	95
— Reply to Comments by C. Burt .. .. .	275
— Observations on Retest Discrepancy in Mentally Sub-normal Children .. .. .	211
VENABLES, E. C. : Placement Problems among Engineering Apprentices in Part-time Technical College Courses. I—Range of Ability .. .. .	237
VENESS, T. : Goal-setting Behaviour, Anxiety and School Streaming .. .. .	23
VERNON, M. D. : The Investigation of Reading Problems To-day .. .. .	146
WILLIAMS, J. L. : Comments on Articles by Mr. D. G. Lewis and Mr. W. R. Jones .. .. .	271



## NOTES TO CONTRIBUTORS

1.—The Editor is glad to consider articles in the fields of educational and child psychology, whether written from the theoretical, clinical, experimental or psychometric standpoints. It is, however, desirable that statistical matter (formulae, tables, discussions of techniques, etc.) should be kept to a minimum, so that the *Journal* will be of interest to statistically untrained readers. Such concepts as the mean, percentiles, standard deviation, correlation, chi-squared, standard error, critical ratio, significance and reliability, and general or group factors, can be employed in the text. But more complex concepts and methods should be explained, or confined to footnotes or an appendix, or published elsewhere.

Articles which have been, or are to be, published elsewhere cannot be accepted.

2.—*Length.* The usual range of length is from 2,500 to 4,500 words, and only very exceptionally will articles of more than 6,000 words be published. Short research reports of 500—1,000 words, such as outlines of theses, are welcomed. These should normally be submitted through the Head of the Department in which the research was carried out.

3.—*Arrangement.* Articles should be preceded by a summary or abstract of between 100 and 300 words, embodying the main conclusions. This should be typed directly after the title and author's name and place of work. A contents analysis, or list of headings, need *not* be included, but successive sections of the article should receive numbered headings, e.g.:

- I. INTRODUCTION.
- II. OUTLINE OF PREVIOUS RESEARCH.
- III. PLAN OF THE INVESTIGATION.
  - (i) *Subjects and materials.*
  - (ii) *Experimental Procedure.*
- IV. STATEMENT OF RESULTS.
- V. DISCUSSION AND CONCLUSIONS.
- VI. REFERENCES.

4.—*Typing.* All papers must be typed (double or 1½-spaced) on one side of quarto or foolscap paper. There should be margins of at least one inch all round. Authors should consult articles in current numbers of the *Journal*, and copy their conventions in respect of capitalising the title and main section headings, italicising author's place of work and sub-section headings, indentation of paragraphs, the form of references and tables, etc.

5.—*Footnotes* should be few in number, and should be indicated in the text by using the signs \*, †, ‡, §.

6.—*References* should normally be indicated by date following the author's name in the text, e.g., Spearman (1904), or (Spearman, 1904, p. 160), or—where there are two or more publications in the same year—Burt (1939a). However, it is occasionally more convenient to number the references throughout, and to type numbers in the text above the line, e.g., Spearman.<sup>4</sup> When there are two or three references only, these may be placed in footnotes in the text, instead of in a separate section at the end.

The list of references at the end should always be in alphabetical order of author's names and—for any one author—by order of date. The following standard form of references should be observed:

DEARBORN, W. F., and ROTHNEY, J. W. M. (1941). *Predicting the Child's Development*. Cambridge, Mass.: Sci.-Art.

(Note that Ltd., etc., after publisher's names, also the numbers of pages in the book, can be omitted.)  
BURT, C. (1939a). The relations of educational abilities. *Brit. J. Educ. Psychol.*, 9, 55-71.

(Note the absence of inverted commas and of capital initials in the titles of articles. The number of the volume is in arabic, not roman numerals. The part of the journal is not given, but page numbers are included in full).

The World List of abbreviations (as given in the December issues of *Psychological Abstracts*) is used for journals.

7.—*Tables and figures* are costly to print, and should be used only when they save space and/or are essential for effective presentation. The same material should not be presented in a table and a graph. Each table or figure should be numbered (in arabic figures), on a separate sheet, and its approximate position indicated in the typescript. Drawings should be in Indian ink on heavy unruled paper, the lettering on graphs being in pencil. Lines in tables should be ruled in pencil. Each table or graph should have a short title, sufficient for understanding without searching in the text.

8.—*Corrections.* The Editor cannot consider for publication articles which are seriously deficient in presentation. The typescript submitted (after preliminary consultation, if desired) should represent the final form in which the author wishes the paper to appear. Since changes in proof are costly, out of all proportion to the original setting, authors must be charged for any changes, insertions or deletions (other than printer's errors).

9.—*Reprints.* Authors receive twenty-five copies of their papers free. Extra copies, in multiples of twenty-five, may be had at cost price if the order is given when the proof is returned.



HOWARD JONES

*Reluctant Rebels*

RE-EDUCATION AND GROUP PROCESS IN A RESIDENTIAL COMMUNITY

A study deriving from the author's work in a residential school for problem children.

Autumn 1960

30s.

P. M. PICKARD

*I Could a Tale Unfold*

THE NATURE OF STORIES FOR CHILDREN

With a Preface by ERNEST JONES

The author traces the history of stories for children from pre-literate forms to the contemporary comic, and has related them to the psychological development of the child as seen in his games, his dreams, and his imaginative life.

Autumn

30s.

J. M. TANNER and BARBEL INHELDER (Editor)

*Discussions on Child Development*

A CONSIDERATION OF THE BIOLOGICAL, PSYCHOLOGICAL, AND CULTURAL APPROACHES TO THE UNDERSTANDING OF HUMAN DEVELOPMENT AND BEHAVIOUR  
VOLUME IV

(Volumes 1-3 were published in 1956, 1956 and 1958 respectively.)

30s.

IVY BENNETT

*Delinquent and Neurotic Children*

A COMPARATIVE STUDY

The aim of this investigation is to compare the character and case histories of delinquent and neurotic children with a view to determining some of the characteristics of their emotional development, and to assess the conditions under which delinquent or neurotic behaviour occurs.

28 October

45s.

MAX WERTHEIMER

*Productive Thinking*

Edited by MICHAEL WERTHEIMER

*Productive Thinking* has become a classic in the fields of psychology and education and is recognised as the most authoritative interpretation of the gestalt view of cognition.

To be published shortly

About 30s.

OSCAR A. OESER

*Teacher, Pupil and Task*

ELEMENTS OF SOCIAL PSYCHOLOGY APPLIED TO EDUCATION

Written for teachers by 14 practising teachers who are also trained psychologists. It makes available some of the modern techniques of handling small groups and provides some of the basic theory of motivation which underlies the learning (and teaching) process.

Second edition

21s.



TAVISTOCK  
PUBLICATIONS





21 JAN 1966